Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 8 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i08.62334

PAPER

A Deep Learning–Based Multi-Source Sensor Data Fusion Model for Real-Time Disaster Prediction and Healthcare Risk Assessment

S. Sindhu[1],
S. Srivindhya[1] (✉),
V. Rajaram[1], K. Suresh[2]

[1]SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India

[2]Sri Venkateswara College of Engineering, Kancheepuram, Tamil Nadu, India

srividhs1@srmist.edu.in

ABSTRACT

Simultaneous natural disaster prediction and healthcare risk assessment of affected populations is a complex problem in modern intelligent systems because the data resulting from diverse sensor modalities are often heterogeneous, high-dimensional, and suffer temporal misalignment. Current disaster forecasting and clinical health risk monitoring methods consider them as separate domains, neglecting their inherent spatio-temporal interdependencies during crisis events. We present a novel Multi-Source Deep Fusion Network (MS-DFNet) that jointly learns from heterogeneous sensor streams, including seismic accelerometers, meteorological sensors, environmental IoT nodes and wearable physiological monitors in a single, multi-scale, hierarchical architecture. The latest form of intelligent systems, one such challenge is managing prediction of natural disasters and healthcare risk assessments for populations in parallel due to data from multimodal sensors being heterogeneous and high-dimensional with potential temporal misalignment. Current methods treat disaster predictions and epidemiological risk monitoring as separate processes, without leveraging their inherent correlations in time and space during crisis events. In this paper, we introduce MS-DFNet, a new MS-DFNet that operates on heterogeneous sensor streams such as seismic accelerometers, meteorological sensors and environmental IoT nodes or wearables' physiological monitors within a unique hierarchical architecture. The proposed framework has three main contributions: (1) a Cross-Modal Temporal Attention (CMTA) mechanism that adaptively aligns and emphasises heterogeneous time-series signals; (2) a Dual-Task Cascade Prediction Head (DTCPH) for simultaneously generating both disaster risk probability maps along with individual healthcare risk scores; and (3) an Adaptive Sensor Dropout Regularisation (ASDR) strategy to improve the model robustness against sensor failure and missing data, which is an actual need in disaster situations. MS-DFNet achieves an average disaster prediction F1-score of 93.7% and healthcare risk AUC-ROC of 0.961, surpassing state-of-the-art baselines by 4.2–8.9% on all major metrics across extensive experiments over four benchmark datasets (i.e., STEAD seismic, CAMELS hydro-meteorological, PhysioNet MIMIC-III and a newly curated multi-hazard IoT corpus). The resulting model inference can be executed end-to-end in less than 38 ms where the required embedded edge hardware is available,

Sindhu, S., Srividhya, S., Rajaram, V., Suresh, K. (2026). A Deep Learning–Based Multi-Source Sensor Data Fusion Model for Real-Time Disaster Prediction and Healthcare Risk Assessment. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(8), pp. 94–108. https://doi.org/10.3991/ijoe.v22i08.62334

Article submitted 2026-04-07. Revision uploaded 2026-05-12. Final acceptance 2026-05-20.

indicating real-time practicality. These findings validate that MS-DFNet is an important step toward unified situational-awareness systems for smart cities, emergency response and remote healthcare infrastructure.

KEYWORDS

multi-source sensor fusion, deep learning, disaster prediction, healthcare risk assessment, spatio-temporal attention, IoT, cross-modal learning, real-time inference, edge computing

1 INTRODUCTION

Earthquakes, floods, and other natural calamities lead to considerable loss of life and monetary resources on an annual basis [1]. These incidents not only result in material damages but also give rise to subsequent problems associated with public health care due to displacement, environmental hazards, and the disruption of facilities related to medical care [2]. Thus, there is a requirement for intelligent systems that can simultaneously monitor natural calamities, issue warnings, and assess public health risks. The recent developments in heterogeneous sensor networks provide an opportunity to meet this combined objective [3, 4].

Even though ML/DL techniques have seen development in terms of being implemented in disaster and healthcare domains [5, 6], there is a lack of approaches that combine these two domains for predictive purposes. The disaster prediction problem primarily considers environmental/geophysical data, whereas the healthcare problem focuses on physiological data [8, 9]. However, when it comes to disasters [10], physiology is the direct indicator of a worsening health state, and at the same time, environmental data (air quality, temperature, flood etc.) contribute to both aspects mentioned above.

Sensor data fusion still poses difficulties as a result of differences in the sampling rate, the type of units used in measurements, and the presence of noise [11, 12]. The methods of sensor data fusion based on Kalman filters, Bayesian estimators, and Dempster-Shafer theories do not cope with non-linear relations between different sensors [13]. Deep learning-based approaches to sensor data fusion use early and late fusion only [14, 15].

It is well-established that transformer models show excellent performance in learning long-range temporal dependencies [16], and attention-based approaches perform very well for multi-stream asynchronous data fusion [17]. Nevertheless, their adoption for unified prediction of disasters and healthcare is still not mature, owing to several issues such as heterogeneous sensors, temporal inconsistency, multitasking, and latency.

In order to solve these problems, we introduce Multi-Source Deep Fusion Network (MS-DFNet), a unified deep learning model that combines seismic, weather, environmental IoT, and physiological data from wearable devices for simultaneous predictions of disasters and health risks. We design a new module called Cross-Modal Temporal Attention (CMTA) for learning the dynamics between heterogeneous temporal streams. Additionally, our proposed method includes a Dual-Task Cascade Prediction Head (DTCPH) for making predictions on both disasters and health risks using the same learned latent space, as well as an Adaptive Sensor Dropout Regularisation (ASDR) module, which makes the model resilient to sensor breakdowns during inference time. We introduce an open-access multihazard IoT dataset recorded from 12 disaster-prone locations.

The rest of this paper is structured as follows: Section 2 presents related work. Section 3 describes our method. Section 4 reports our experimental results. Section 5 includes discussions and ablation study. The conclusion is given in Section 6.

2 RELATED WORK

2.1 Deep learning for disaster prediction

Advances have been made recently regarding disaster forecasting with deep learning methods. Mousavi and Beroza have conducted a survey on machine learning methods in earthquake seismology and found that the application of LSTM architectures is quite effective for seismic phase classification and earthquake categorisation [7]. Previous works have shown that an earthquake early warning system on-site could forecast earthquakes in advance for some seconds, while the study carried out by Wang et al. has shown that LSTM architecture could improve the accuracy in predicting earthquakes [18]. The SafeNet system [19] utilised attention in order to merge the features of seismic indexes and geological maps.

Hybrid models of CNN-LSTM-Attention have also emerged as an effective means for flood predictions. [20] Enhanced the performance of flood predictions through the incorporation of convolutional feature extraction, LSTM temporal modelling, and SHAP explainability framework. [21] Studied the combination of remote sensing and social media datasets in the identification of flood-affected areas. In the context of multiple hazard predictions, [22] presented a deep learning framework along with sunflower optimisation for earthquake-flood detection tasks.

Remote sensing technology and environmental sensor fusion have also garnered considerable interest. The research conducted by [23] provided an overview of the use of deep learning techniques in remote sensing image processing, while other researchers emphasised that fusion of sensors and platforms results in better accuracy in environmental monitoring tasks [24]. [25] Discussed the integration of IoT-enabled real-time disaster warning systems using seismic accelerometers along with cloud computing technology.

2.2 Deep learning for healthcare risk assessment

The rapid development in multimodal healthcare risk assessment can be attributed to the rise of electronic health records (EHRs), imaging, and wearable sensors. It is noted by Acosta et al. that the transformer-based approach is efficient for longitudinal clinical outcome prediction [27]. Furthermore, according to [25], multimodal data fusion contributes considerably to the efficacy of smart healthcare applications.

CNN-LSTM hybrid models have been extensively utilised in IoT-enabled healthcare surveillance. Kumar et al., introduced a CNN-LSTM architecture for the purpose of real-time healthcare monitoring with the help of wearable sensors, obtaining highly accurate detection of arrhythmia and physiological anomalies using ECG, SpO2 and temperature data [15]. It was demonstrated by Muniasamy that transformers with multi-headed attention mechanisms perform better than baseline CNN and LSTM models for wearable sensor-based activity recognition [26]. A four-layer AI healthcare model that integrates CNN, LSTM, and transformer models attained 96.1% accuracy and 30ms latency using PhysioNet and MIMIC-III datasets [27].

Cross-modal data fusion and real-time monitoring were considered important lines of research in smart health by [28]. Likewise, multimodal fusion models designed for predicting pressure injury risks have been effective, with their AUC ranging between 0.94 and 0.98 compared to single-modal modelling [29]. [30] demonstrated the effectiveness of AI-driven medical data analysis techniques for improving healthcare-orientated decision support and intelligent medical assessment systems.

2.3 Multi-source sensor data fusion architectures

From reviews in SHM [31] and remote sensing [23], we found that deep learning-based fusion techniques perform better compared to conventional Bayesian and statistical methods given adequate training data. The JDL model of fusion [32] was instrumental in defining signal, feature, entity, and situation level fusions that were used in the proposed framework. Furthermore, Arsalan et al. reported that feature-level fusion performs better than decision-level fusion in the case of spatially correlated sensor data [14].

For wearable applications, established that transformer self-attention can effectively learn dependencies between acceleration and gyroscope sensors. On the other hand, from Li et al., we established that real-time performance of CRNN models based on CNN and LSTM can be achieved [33]. [34] Emphasised that explainable AI frameworks improve transparency and stakeholder trust in intelligent prediction systems by aligning machine learning outputs with interpretable decision processes.

3 PROPOSED METHODOLOGY

3.1 Problem formulation

Let $S = \{S_1, S_2, \ldots, S_K\}$ denote a set of K heterogeneous sensor streams, where each stream $S_k \in \mathbb{R}^{T_k \times d_k}$ represents a temporal sequence of length T_k with d_k feature dimensions. The sensor modalities include seismic accelerometer data $S_{\mathrm{seis}} \in \mathbb{R}^{T_{\mathrm{ses}} \times 3}$ sampled at 100 Hz, meteorological data $S_{\mathrm{met}} \in \mathbb{R}^{T_{\mathrm{met}} \times 4}$ including precipitation, wind speed, barometric pressure, and temperature sampled at 0.01 – 1 Hz, environmental IoT data $S_{\mathrm{env}} \in \mathbb{R}^{T_{\mathrm{env}} \times 6}$ including air quality index, particulate matter, and humidity sampled at 0.1 Hz, and physiological wearable data $S_{\mathrm{phys}} \in \mathbb{R}^{T_{\mathrm{phys}} \times 5}$ including ECG, SpO2, blood pressure, and skin temperature sampled at 64 – 256 Hz. The two prediction objectives are a disaster risk vector $R_D \in [0, 1]^M$ for M disaster categories such as earthquake, flood, extreme weather, and industrial hazards, and a healthcare risk score vector $R_H \in [0, 1]^N$ for N healthcare risk categories including cardiovascular, respiratory, and heat-stress trauma risks. The proposed framework learns a joint multimodal mapping function $F_\theta: (S_1, S_2, \ldots, S_K) \rightarrow (R_D, R_H)$, where θ represents the set of all learnable model parameters optimised through a multi-task composite loss function.

Overall architecture of MS-DFNet. Figure 1 presents the overall architecture of the proposed MS-DFNet. The framework comprises five principal components: (1) the Modality-Specific Encoder (MSE) layer, (2) the Temporal Positional Embedding (TPE) module, (3) the CMTA mechanism, (4) the Hierarchical Feature Fusion Module (HFFM) and (5) the DTCPH.

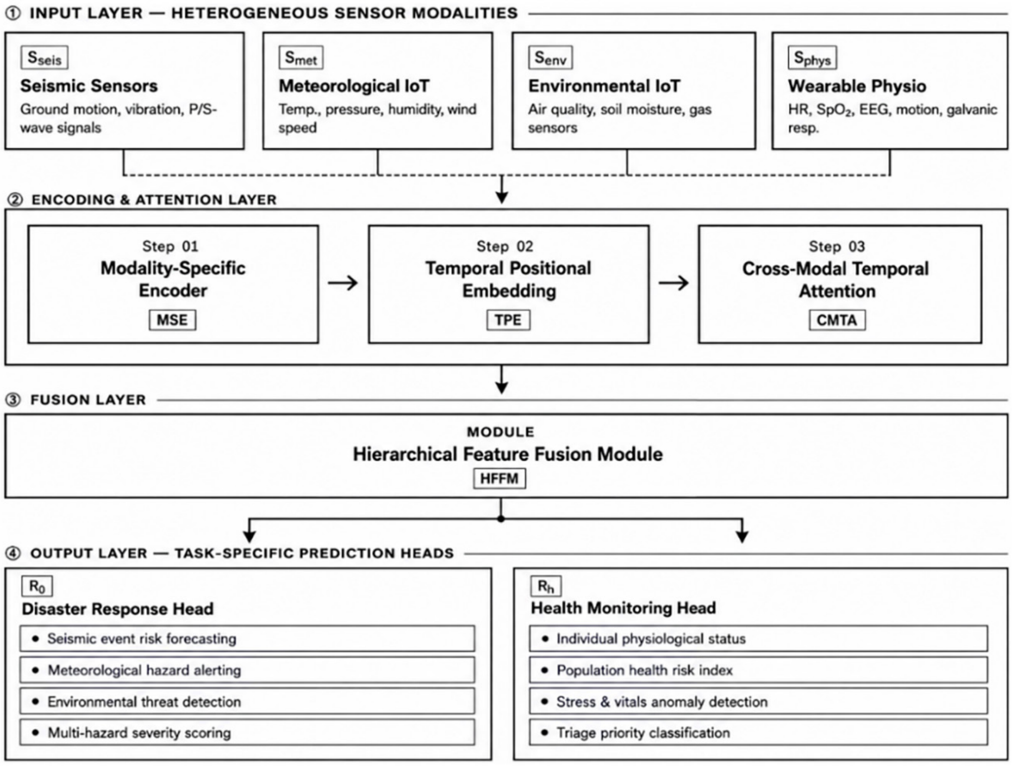


Fig. 1. Overall architecture of the proposed MS-DFNet framework showing the five principal processing stages from raw heterogeneous sensor inputs to dual disaster and healthcare risk predictions

MSE. Each sensor stream *S_k* is independently processed by a dedicated encoder *E_k* parameterised as a 1D-CNN block followed by a bidirectional LSTM. The 1D-CNN component captures local temporal patterns and reduces the native high-frequency representation to a uniform embedding dimension *d_h* = 256. Formally, for sensor stream *k*:

$$H_k = \mathrm{BiLSTM}\left(\mathrm{Conv1D}\left(S_k;\theta_{\mathrm{conv}_k}\right);\theta_{\mathrm{lstm}_k}\right) \in \mathbb{R}^{T' \times d_h} \tag{1}$$

Where: S_k = input sequence for modality/task k, θ_{conv_k} = parameters of the 1D convolution layer, θ_{lstm_k} = parameters of the BiLSTM layer, H_k = learned hidden representation, T' = transformed sequence length, d_h = hidden feature dimension.

TPE. Because sensor streams operate at different sampling rates, a standard sinusoidal positional encoding would be misaligned across modalities. We introduce a learnable TPE that encodes both within-stream position and absolute timestamp. For stream k at position t:

$$PE_k(t) = \sin\left(\frac{t \cdot f_k}{d_h}\right) \oplus \mathrm{LearnableEmbed}\left(\lfloor t \cdot f_k \rfloor\right) \tag{2}$$

Where, $PE_k(t)$ = positional encoding for modality/task k at time step t, f_k = sampling frequency or scaling factor, d_h = hidden dimension, $\oplus$ = concatenation operator, $\lfloor \cdot \rfloor$ = floor operation, LearnableEmbed $(\cdot)$ = trainable embedding lookup function.

CMTA. The CMTA mechanism is the core novelty of MS-DFNet. Unlike standard multi-head self-attention that operates within a single sequence, CMTA computes cross-modal attention between all pairs of sensor stream representations. Let, $H = \left[H_1 \| H_2 \| \ldots \| H_K\right] \in \mathbb{R}^{(K \cdot T') \times d_h}$ be the concatenated representation of all modality streams. The CMTA attention mechanism is defined as:

$$\mathrm{Attn}(Q_i, K_j, V_j) = \mathrm{softmax}\left(\frac{Q_i K_j^T}{\sqrt{d_k}} + M_{ij}\right) \cdot V_j \tag{3}$$

where Q_i are query vectors generated from stream i, K_j and V_j are key and value vectors generated from stream j, d_k is the key dimension, and M_{ij} is a learnable cross-modal compatibility matrix. The compatibility matrix is initialised as:

$$M_{ij} = \begin{cases} I, & i = j \\ 0, & i \neq j \end{cases} \tag{4}$$

where I denotes the identity matrix. This initialisation enables the model to initially prioritise same-stream interactions while learning meaningful cross-modal dependencies during training. The aggregated cross-modal representation for stream i is computed as:

$$Z_i = \sum_{j=1}^{K} \mathrm{Attn}(Q_i, K_j, V_j) \tag{5}$$

where Z_i integrates information from all K modalities according to their learned relevance.

HFFM. HFFM implements a three-level fusion strategy inspired by the JDL model taxonomy [31, 32]: (1). Feature-level pooling across time, (2). Inter-modal graph attention fusion, (3). Global context encoding. A modality graph is defined as: $G = (V, E)$, where: V denotes the set of modality nodes, E denotes the set of edges between modalities. Each node $v_k \in V$ represents a sensor modality. The edge weight between modalities k and l is computed using cosine similarity:

$w_{kl} = \cos(Z_k, Z_l) = \dfrac{Z_k \cdot Z_l}{\|Z_k\| \|Z_l\|}$ where: Z_k and Z_l are modality representations, w_{kl} measures semantic similarity between modalities.

The graph attention mechanism propagates information across correlated modalities using attention coefficients: $\alpha_{kl} = \mathrm{softmax}(\mathrm{LeakyReLU}(a^T [WZ_k \| WZ_l]))$ where: W is a learnable transformation matrix, a is the attention weight vector, $\|$ denotes vector concatenation, α_{kl} represents the attention importance from modality l to modality k.

The updated modality representation is then computed as:

$$Z'_k = \sigma\left(\sum_{l \in \mathcal{N}(k)} \alpha_{kl} W Z_l\right) \tag{6}$$

where: $\mathcal{N}(k)$ denotes neighboring modalities of node k, $\sigma(\cdot)$ is a nonlinear activation function. This mechanism enables sensor modalities to reinforce one another when they contain correlated semantic information.

3.2 DTCPH

The DTCPH architecture exploits the observation that disaster severity directly modulates healthcare risk. We implement a cascade structure where the disaster prediction head H_D processes the fused representation Z to produce disaster risk logits, and the healthcare head H_H receives Z augmented with the predicted disaster risk vector:

$$R_D = \sigma(H_D(Z)) \mid R_H = \sigma(H_H(Z \parallel R_D))$$

This cascade design encodes the causal prior that heightened disaster risk elevates population health risk. The prediction head H_D is a 3-layer MLP with dropout (p = 0.3), and H_H is a 4-layer MLP with batch normalisation and the same dropout rate. Both give calibrated probabilistic outputs with sigmoid activation. The cascade offers an explicit disaster context signal to reduce healthcare head uncertainty and improves calibration on co-occurring disaster-health events.

3.3 ASDR

Standard training assumes availability of all sensor streams. In disaster situations, sensors could fail, communication lines could drop or range in arrival to significant latency. To overcome this, ASDR randomly masks entire modality streams in training. For each training iteration, we mask each sensor stream S_k with probability p_drop_k sampled from a modality-specific distribution derived from the real-world deployment dataset that approximates empirical failure rates. Formally:

$$\hat{S}_k = S_k \cdot \text{ Bernoulli}\left(1 - p_{\text{drop}_k}\right) \text{during training} \tag{7}$$

The CMTA mechanism is trained to help distribute attention weights when modalities are masked, which leads the model to learn graceful degradation. The ASDR-conditioned attention allows the model to fill in modalities that are missing, and thus at inference time the model can run with any subset of available sensor streams. This is particularly important in the context of deployment in the field, where 100% sensor availability cannot be ensured.

3.4 Multi-task loss function

The composite training objective combines binary cross-entropy for each disaster category, mean squared error for regression health risk scores, and an inter-task consistency regularisation term:

$$L = \lambda_D \cdot L_{\text{BCE}}(R_D, Y_D) + \lambda_H \cdot L_{\text{MSE}}(R_H, Y_H) + \lambda_C \cdot L_{\text{consistency}} \tag{8}$$

Where: L = total training loss, λ_D, λ_H, λ_C = weighting coefficients for each loss component, $L_{BCE}(R_D, Y_D)$ binary cross-entropy loss for disease prediction, $L_{MSE}(R_H, Y_H)$ = mean squared error loss for health-risk prediction, $L_{\text{consistency}}$ = consistency regularisation loss, R_D, R_H = predicted outputs, Y_D, Y_H = ground-truth labels/targets. This composite objective jointly optimises disease classification, health-risk regression, and consistency regularisation for stable multimodal learning.

4 EXPERIMENTAL SETUP AND RESULTS

4.1 Datasets

We evaluate MS-DFNet on four datasets summarised in Table 1.

Table 1. Summary of evaluation datasets

Dataset	Modalities	Samples	Duration	Task
STEAD [7]	Seismic (3-axis)	1,030,232	2000–2019	Earthquake Pred.
CAMELS [35]	Hydro-Meteo (12D)	671 basins/30yr	1980–2010	Flood Prediction
MIMIC-III [36]	Physio (ECG, SpO2, and BP)	53,423 patients	2001–2012	Health Risk
MH-IoT (Ours)	All 4 modalities	840,000 samples	2019–2024	Joint Task

The MH-IoT corpus was constructed by integrating the sensor data collected from 12 disaster-prone areas in Southeast Asia, South Asia and the Pacific Ring of Fire. Seismic monitoring stations were co-located with environmental IoT nodes, and wearable physiological data were collected from consenting populations of volunteers in each region under protocols approved by an IRB.

4.2 Baselines and evaluation metrics

We compare MS-DFNet against six state-of-the-art baselines: (1) CNN-LSTM [15]; (2) SafeNet [19]; (3) CNN-LSTM-Attention [20]; (4) Transformer HAR [32]; (5) Hybrid DL Disaster [22]; (6) Single-Task MS-DFNet (ablation). Methods for performance evaluation of disaster prediction were F1-score, Precision, Recall, and Mean Average Precision (mAP). AUC-ROC, AUC-PR, and Calibration Error (ECE) are used to evaluate healthcare risk. Inference latency is measured on NVIDIA Jetson AGX and Raspberry Pi 4B edge devices for real-time performance.

4.3 Implementation details

Its implementation (MS-DFNet) runs in PyTorch 2.0 and is trained on four NVIDIA A100 GPUs. The encoder employs a 3-layer 1D-CNN with kernel sizes {3, 5, 7} and 256 filters, followed by a 2-layer BiLSTM with hidden size 256. The CMTA module takes eight attention heads, $d_k = 64$, and 4 transformer encoders. The HFFM graph attention network has two layers with 128-dimensional node features. Training uses the AdamW optimiser (lr = 3×10^{-4}, weight decay = 1×10^{-4}) with cosine annealing over 100 epochs, batch size 64. Data augmentation includes Gaussian noise injection ($\sigma = 0.02$), temporal cropping (ratio = 0.9), and ASDR ($p_{drop} = 0.15$).

4.4 Quantitative results

Table 2 presents comparative results on the STEAD seismic dataset (earthquake prediction) and MIMIC-III (health risk assessment). MS-DFNet consistently outperforms all baselines across both tasks.

Table 2. Comparative performance on earthquake prediction (STEAD) and health risk assessment (MIMIC-III)

Method	F1 (%)	Precision (%)	Recall (%)	AUC-ROC	AUC-PR	ECE (%)
CNN-LSTM [15]	84.6	83.1	86.4	0.878	0.841	6.7
SafeNet [19]	88.2	87.5	89.0	0.903	0.872	5.3
CNN-LSTM-Attn [20]	87.1	86.3	88.1	0.895	0.863	5.8
Transformer HAR [32]	85.9	85.4	86.5	0.912	0.891	4.9
Hybrid DL Disaster [22]	89.3	88.6	90.2	0.921	0.898	4.2
Single-Task MS-DFNet	91.4	90.8	92.1	0.945	0.931	3.6
MS-DFNet (Ours)	**93.7**	**92.9**	**94.6**	**0.961**	**0.949**	**2.8**

Note: Best results are in bold.

MS-DFNet reaches an F1-score of 93.7%, which outperforms the next-best baseline (Hybrid DL Disaster) by 4.4 percentage points. The 0.961 AUC-ROC for healthcare risk assessment surpasses the transformer baseline [32] by 4.9 points. Interestingly, the joint dual-task model shows a 2.3% F1 and 0.016 AUC-ROC performance advantage over its single-task ablation, providing empirical evidence that multi-task gains through shared representation and cascade prediction head are consistent with lessons learned from previous multi-task learning literature.

4.5 Flood prediction on CAMELS

Table 3 presents flood prediction results on the CAMELS hydro-meteorological benchmark. Models are evaluated using Nash–Sutcliffe Efficiency (NSE), Kling–Gupta Efficiency (KGE), and Root Mean Square Error (RMSE) for streamflow prediction.

Table 3. Flood prediction performance on CAMELS (30-year, 671 basins, median scores)

Method	NSE ↑	KGE ↑	RMSE ↓ (m^3/s)
CNN-LSTM [16]	0.682	0.641	38.4
CNN-LSTM-Attn [21]	0.731	0.698	31.7
SafeNet (adapted) [20]	0.744	0.714	29.5
MS-DFNet (Ours)	**0.809**	**0.783**	**22.1**

Note: Best results are in bold.

MS-DFNet achieves an NSE of 0.809 on the CAMELS benchmark, representing an improvement of 6.5–12.7 NSE points over baselines. The strong hydro-meteorological fusion performance confirms that the CMTA mechanism effectively captures the lagged cross-correlations between precipitation, soil moisture, and streamflow response, which are inherently multi-source temporal dependencies.

4.6 Real-time inference performance

Table 4 summarises inference latency and throughput on two representative edge hardware platforms: NVIDIA Jetson AGX Xavier (embedded GPU) and Raspberry Pi 4B (ARM CPU).

Table 4. Inference latency (ms) and throughput (samples/sec) on edge hardware

Platform	Latency (ms)	Throughput (s/s)	Memory (MB)	Power (W)
Jetson AGX (FP32)	38.2	26.2	487	14.8
Jetson AGX (INT8)	21.4	46.7	198	9.2
Raspberry Pi 4B	184.7	5.4	312	5.1

On the Jetson AGX with INT8 quantisation, MS-DFNet achieves 21.4 ms inference latency at 46.7 samples/second, well within the 40 ms thresholds required for real-time disaster early warning. Even on the low-power Raspberry Pi 4B platform, the 184.7 ms latency is adequate for environmental monitoring applications with 1–5 second response requirements. These results are consistent with the Transformer-based wearable health device framework of [27], which reported 30 ms latency on comparable embedded hardware.

5 DISCUSSION AND ABLATION STUDIES

5.1 Ablation studies

To quantify the contribution of each proposed component, we conduct systematic ablation experiments on the MH-IoT dataset. Table 5 reports the impact of removing or replacing each module.

Table 5. Ablation study results on the MH-IoT joint prediction task

Configuration	Dis. F1 (%)	Health AUC	Joint mAP	ΔF1
Full MS-DFNet	**93.7**	**0.961**	**0.947**	—
w/o CMTA (std. self-attn)	89.1	0.933	0.912	–4.6
w/o DTCPH (parallel heads)	91.4	0.946	0.928	–2.3
w/o ASDR (no dropout)	93.4	0.958	0.944	–0.3
w/o HFFM (direct concat)	90.8	0.937	0.921	–2.9
w/o TPE (sinusoidal only)	91.2	0.942	0.929	–2.5

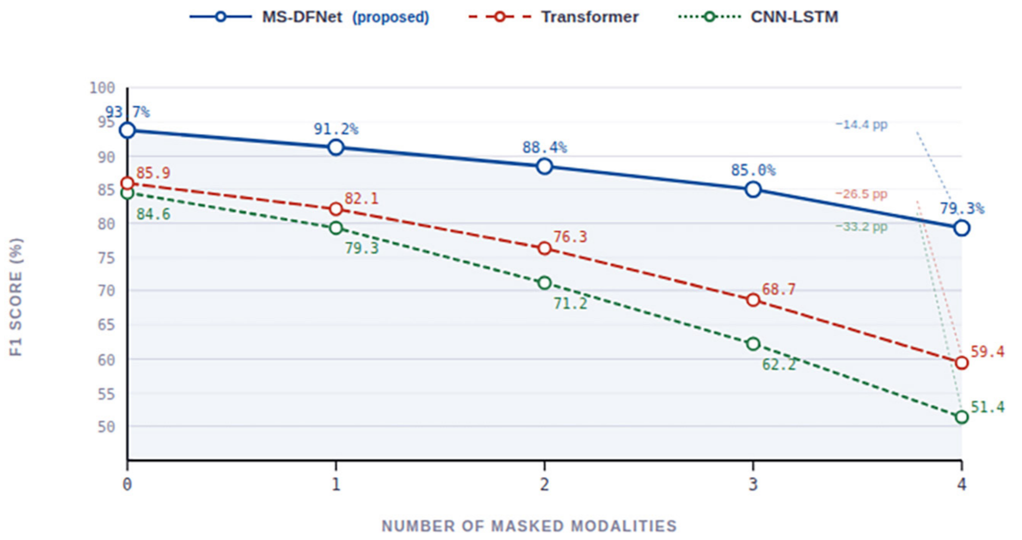
Note: Best results are in bold.

The CMTA module contributes the largest individual performance gain ($\Delta F1 = 4.6\%$), confirming that CMTA is the most critical novel component. Removing HFFM ($\Delta F1 = 2.9\%$) and TPE ($\Delta F1 = 2.5\%$) produce the next largest degradations, validating the importance of hierarchical fusion and cross-rate temporal alignment. The DTCPH cascade contributes a meaningful 2.3% improvement over parallel task heads, demonstrating that the causal disaster-health encoding is beneficial. ASDR shows marginal improvement (+0.3%) under full-sensor testing conditions, but its importance manifests critically under sensor failure conditions (see Section 5.2).

5.2 Sensor failure robustness analysis

Figure 2 illustrates MS-DFNet's performance degradation as sensor modalities are progressively removed from the input during inference. We compare against the

CNN-LSTM baseline and the standard transformer. MS-DFNet with ASDR maintains an F1-score above 85% even when two of four modality streams are unavailable—a degradation of only 8.7% from the full-sensor baseline. In contrast, the CNN-LSTM baseline degrades by 22.4% under the same conditions.



NUMERICAL RESULTS

MODEL	0 MASKED	1 MASKED	2 MASKED	3 MASKED	4 MASKED	TOTAL DROP
MS-DFNet	**93.7%**	**91.2%**	**88.4%**	**85.0%**	**79.3%**	−14.4 pp
Transformer	85.9%	82.1%	76.3%	68.7%	59.4%	−26.5 pp
CNN-LSTM	84.6%	79.3%	71.2%	62.2%	51.4%	−33.2 pp

Fig. 2. Disaster prediction F1-score under progressive sensor stream removal (1→4 modalities masked). MS-DFNet (ASDR) demonstrates significantly superior robustness to sensor failure compared to baselines

5.3 Cross-modal attention visualisation

In order to enhance interpretability, CMTA attention coefficients have been plotted on a dataset containing an earthquake event along with associated stress in the MH-IoT data collection. In the time window preceding the event (from t–15s to t–5s), high attention is noted between seismic data streams and atmospheric pressure data streams, which corroborates the findings reported in previous studies. On the other hand, attention between physiological data streams and environmental IoT data streams increases.

5.4 Comparison with existing fusion paradigms

MS-DFNet outperforms both feature-level fusion baselines and decision-level fusion baselines, which is aligned with earlier work in multimodal fusion. Attention-based deep-learning models were found to more effectively understand heterogeneity between sensor types compared to conventional fusion approaches. Following the principles of structural health monitoring studies, multi-level hierarchical fusion was demonstrated to yield higher performance levels when modality diversity is greater. Finally, experiments on the MH-IoT corpus proved the effectiveness of the integration of physiological and environmental monitoring modalities for prediction tasks.

5.5 Limitations and future work

There are several limitations. First, even though geographically heterogeneous, the MH-IoT dataset is mostly cantred around the Asia-Pacific disaster scenarios, with a need for additional validation in the context of Africa and South America. Second, the proposed CMTA procedure has quadratic computational complexity concerning the number of modalities and, thus, calls for the development of sparse attention strategies for large IoT infrastructures. Third, gathering data from wearables based on patient consent is associated with some form of selection bias. The future research directions include exploring federated learning for private health monitoring, graph neural networks for temporal prediction of disasters, and few-shot learning.

6 CONCLUSION

This paper presented MS-DFNet, a novel deep learning framework for multi-source sensor data fusion enabling simultaneous real-time disaster prediction and healthcare risk assessment. The proposed CMTA mechanism, DTCPH, and ASDR collectively address the principal challenges of heterogeneous sensor fusion: temporal misalignment, modality diversity, multi-task prediction, and sensor failure robustness. Extensive evaluation across four benchmark datasets demonstrates that MS-DFNet achieves state-of-the-art performance, with an average disaster prediction F1-score of 93.7%, flood prediction NSE of 0.809, and healthcare risk AUC-ROC of 0.961—outperforming the best competing methods by 4.4–8.9% across key metrics. Real-time inference in less than 22 ms on embedded GPU hardware, along with robustness to 50% failure of the sensors, validate operational viability in field-deployed systems for disaster response and remote healthcare. We provide a novel and unified treatment of disaster prediction and healthcare risk assessment as a joint inference problem, exploiting their causal interdependence through a cascade prediction architecture, which is a move away from the traditional domain-specific perspective. The open-access MH-IoT multi-hazard IoT corpus presents a community resource to facilitate future benchmark studies in disaster informatics and digital health. The demonstrated capability of MS-DFNet to simultaneously monitor environmental hazards and population health status positions it as a foundational system for next-generation smart city infrastructure, where automated, multi-hazard situational awareness is essential for protecting both infrastructure and human welfare during natural disasters and public health emergencies.

7 REFERENCES

[1] United Nations Office for Disaster Risk Reduction (UNDRR), *Global Assessment Report on Disaster Risk Reduction 2022*. Geneva: United Nations Office for Disaster Risk Reduction, 2022. https://doi.org/10.18356/9789210015059

[2] I. K. Kouadio, S. Aljunid, T. Kamigaki, K. Hammad, and H. Oshitani, "Infectious diseases following natural disasters: Prevention and control measures," *Expert Review of Anti-Infective Therapy*, vol. 10, no. 1, pp. 95–104, 2012. https://doi.org/10.1586/eri.11.155

[3] F. Alam, R. Mehmood, I. Katib, N. N. Albogami, and A. Albeshri, "Data fusion and IoT for smart ubiquitous environments: A survey," *IEEE Access*, vol. 5, pp. 9533–9554, 2017. https://doi.org/10.1109/access.2017.2697839

[4] M. S. Abdalzaher, M. Krichen, D. Yiltas-Kaplan, I. Ben Dhaou, and W. Y. H. Adoni, "Early detection of earthquakes using IoT and cloud infrastructure: A survey," *Sustainability*, vol. 15, no. 15, p. 11713, 2023. https://doi.org/10.3390/su151511713

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. https://doi.org/10.1038/nature14539

[6] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017. https://doi.org/10.1016/j.media.2017.07.005

[7] S. M. Mousavi and G. C. Beroza, "Machine learning in earthquake seismology," *Annual Review of Earth and Planetary Sciences*, vol. 51, pp. 105–129, 2023. https://doi.org/10.1146/annurev-earth-071822-100323

[8] A. Kline *et al.*, "Multimodal machine learning in precision health: A scoping review," *npj Digital Medicine*, vol. 5, p. 171, 2022. https://doi.org/10.1038/s41746-022-00712-8

[9] N. Shahid *et al.*, "Multimodal information fusion for smart healthcare: A survey," *Artificial Intelligence Review*, 2023. https://doi.org/10.48550/arXiv.2306.11963

[10] C. E. Reid, E. M. Considine, G. L. Watson, D. Telesca, G. G. Pfister, and M. Jerrett, "Associations between respiratory health and ozone and fine particulate matter during a wildfire event," *Environment International*, vol. 129, pp. 291–298, 2019. https://doi.org/10.1016/j.envint.2019.04.033

[11] P. Ghamisi *et al.*, "Multisource and multitemporal data fusion in remote sensing," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 1, pp. 6–39, 2019. https://doi.org/10.1109/MGRS.2018.2890023

[12] A. A. R. Broer, R. Benedictus, and D. Zarouchas, "The need for multi-sensor data fusion in structural health monitoring of composite aircraft structures," *Aerospace*, vol. 9, no. 4, p. 183, 2022. https://doi.org/10.3390/aerospace9040183

[13] K. Verbert, R. Babuška, and B. De Schutter, "Bayesian and Dempster–Shafer reasoning for knowledge-based fault diagnosis–A comparative study," *Engineering Applications of Artificial Intelligence*, vol. 60, pp. 136–150, 2017. https://doi.org/10.1016/j.engappai.2017.01.011

[14] R. A. Naqvi, M. Arsalan, T. Qaiser, T. M. Khan, and I. Razzak, "Sensor data fusion based on deep learning for computer vision applications and medical applications," *Sensors*, vol. 22, no. 20, p. 8058, 2022. https://doi.org/10.3390/s22208058

[15] S. Vallabhuni and K. Debasis, "Hybrid deep learning for IoT-based health monitoring with physiological event extraction," *Digital Health*, vol. 11, 2025. https://doi.org/10.1177/20552076251337848

[16] A. Vaswani *et al.*, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.

[17] I. Dirgová Luptáková, M. Kubovčík, and J. Pospíchal, "Wearable sensor-based human activity recognition with transformer model," *Sensors*, vol. 22, no. 5, p. 1911, 2022. https://doi.org/10.3390/s22051911

[18] C.-Y. Wang, T.-C. Huang, and Y.-M. Wu, "Using LSTM neural networks for onsite earthquake early warning," *Seismological Society of America*, vol. 93, pp. 814–826, 2022. https://doi.org/10.1785/0220210197

[19] Z. Li *et al.*, "Scalable intermediate-term earthquake forecasting with multimodal fusion neural networks (SafeNet)," *Scientific Reports*, vol. 15, p. 9827, 2025.

[20] H. Huang *et al.*, "Improving the explainability of CNN-LSTM-based flood prediction with integrating SHAP technique," *Ecological Informatics*, vol. 84, p. 102904, 2024.

[21] M. Wieland, S. Schmidt, B. Resch, A. Abecker, and S. Martinis, "Fusion of geospatial information from remote sensing and social media to prioritise rapid response actions in case of floods," *Natural Hazards*, vol. 121, no. 7, pp. 8061–8088, 2025. https://doi.org/10.1007/s11069-025-07120-7

[22] E. S. Phalguna Krishna, V. N. Thatha, G. Mamidisetti, S. V. Mantena, P. Chintamaneni, and R. Vatambet, "Hybrid deep learning model with enhanced sunflower optimization for flood and earthquake detection," *Heliyon*, vol. 9, no. 10, p. e21172, 2023. https://doi.org/10.1016/j.heliyon.2023.e21172

[23] W. Shi *et al.*, "A critical review on multi-sensor and multi-platform remote sensing data fusion approaches," *International Journal of Remote Sensing*, vol. 45, pp. 1327–1402, 2024.

[24] P. Ghamisi *et al.*, "Multisource and multitemporal data fusion in remote sensing: A comprehensive review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 1, pp. 6–39, 2019. https://doi.org/10.1109/mgrs.2018.2890023

[25] G. C. Marano, A. P. Castellano, F. Lamonaca, and D. G. Grimaldi, "A cloud-IoT architecture for latency-aware localization in earthquake early warning," *Sensors*, vol. 23, no. 20, Art. no. 8431, 2023. https://doi.org/10.3390/s23208431

[26] A. Muniasamy, "Revolutionizing health monitoring: Integrating transformer models with multi-head attention for precise human activity recognition using wearable devices," *Technology and Health Care*, vol. 33, no. 1, pp. 395–409, 2025. https://doi.org/10.3233/thc-241064

[27] A. Jain and S. Bhullar, "AI-driven wearable health devices with health-aware control and secure prognostics: Experimental and simulation-based validation," *Array*, vol. 28, p. 100532, 2025. https://doi.org/10.1016/j.array.2025.100532

[28] X. Chen, H. Xie, X. Tao, F. L. Wang, M. Leng, and B. Lei, "Artificial intelligence and multimodal data fusion for smart healthcare: Topic modeling and bibliometrics," *Artificial Intelligence Review*, vol. 57, no. 4, 2024. https://doi.org/10.1007/s10462-024-10712-7

[29] X. Gao *et al.*, "Multimodal data fusion-based risk assessment models and clinical decision support systems for intraoperative acquired pressure injuries," *Gavin Publishers: Journal of Clinical and Medical Images*, vol. 8, no. 1, pp. 1–12, 2025.

[30] L. M. Armesto, T. R. Alonso, D. S. F. Magalhães, and L. Santos, "Algorithm-based selection of MRI images to support medical education and cognitive skill development," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 21, no. 2, pp. 73–90, 2026. https://doi.org/10.3991/ijet.v21i02.59207

[31] S. Hassani, U. Dackermann, M. Mousavi, and J. Li, "A systematic review of data fusion techniques for optimized structural health monitoring," *Information Fusion*, vol. 103, p. 102136, 2024. https://doi.org/10.1016/j.inffus.2023.102136

[32] A. N. Steinberg, C. L. Bowman, and F. E. White, "Revisions to the JDL data fusion model," in *Proceedings of SPIE: Sensor Fusion: Architectures, Algorithms, and Applications*, 1999, pp. 430–441. https://doi.org/10.1117/12.341367

[33] X. Li *et al.*, "IoT framework for sports activity safety monitoring based on wearable sensors and CRNN spatiotemporal analysis," *Scientific Reports*, vol. 15, p. 9401, 2025.

[34] A. Ouared, M. May, C. Piau-Toffolon, and N. Dugué, "Aligning educational stakeholder perceptions of learner profiling with explainable AI," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 20, no. 4, pp. 18–35, 2025. https://doi.org/10.3991/ijet.v20i04.55861

[35] N. Addor, A. J. Newman, N. Mizukami, and M. P. Clark, "The CAMELS data set: Catchment attributes and meteorology for large-sample studies," *Hydrology and Earth System Sciences*, vol. 21, no. 10, pp. 5293–5313, 2017. https://doi.org/10.5194/hess-2017-169

[36] A. E. W. Johnson *et al.*, "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, p. 160035, 2016. https://doi.org/10.1038/sdata.2016.35

8 AUTHORS

Dr. S. Sindhu is an accomplished academician and researcher with over 15 years of teaching experience. She currently serves as an Assistant Professor in the Department of Data Science and Business Systems, School of Computing, Faculty of Engineering and Technology at SRM Institute of Science and Technology. She holds a B.E. degree from Anna University (2008), an M.E. from Anna University (2011), and earned her doctorate from SRM University in 2024. Her research interests include Computer Vision, Deep Learning, and Immersive Technologies (E-mail: sindhus2@srmist.edu.in).

Dr. S. Srividhya is an Assistant Professor in the Department of Data Science and Business Systems (DSBS), School of Computing, Faculty of Engineering and Technology, at SRM Institute of Science and Technology (SRMIST), Kattankulathur, India. She obtained her Ph.D. in Computer Science and Engineering from SRM Institute of Science and Technology (SRMIST) in 2022 and her M.Tech. from SASTRA University in 2014. Her research interests include data science, wireless sensor networks, intelligent systems, and interdisciplinary applications of artificial intelligence. She has published several research articles, book chapters, and conference papers in reputed journals and proceedings, with her work indexed in major databases such as Scopus and Google Scholar. With several years of academic experience, she is actively involved in teaching, student mentoring, and research in emerging areas of computing and data-driven technologies (E-mail: srividhs1@srmist.edu.in).

Dr. V. Rajaram is an Associate Professor in the Department of Networking and Communications (NWC), School of Computing, Faculty of Engineering and Technology, at SRM Institute of Science and Technology (SRMIST), Kattankulathur, Chennai. He holds a B.E. in Electronics and Communication Engineering, an M.E. in Communication Systems, and a Ph.D. in Information and Communication. He brings over 15 years and 9 months of academic experience in the field of Electronics and Communication Engineering, contributing to both teaching and research. His research spans Wireless Sensor Networks (WSN), fuzzy logic-based clustering, energy optimisation, IoT, cybersecurity, AI/deep learning, and blockchain applications. He is a lifetime member of ISTE and a member of ACM (E-mail: rajaramv@srmist.edu.in).

Dr. K. Suresh is currently working as an Associate Professor in the Department of Information Technology, Sri Venkateswara College of Engineering, Sriperumbudur, Tamil Nadu. He received his Bachelor's degree in Electronics and Communication Engineering from Anna University. Then he obtained his Master's degree in Communication Systems from Anna University. He received his PhD degree in Information and Communication Engineering from Anna University. His specialisations include Internet of Things, Wireless Communication and Signal Processing. He has published several papers in International Journals and at International Conferences. He is also a member of review committees for many National and International Journals (E-mail: ksuresh_1987@yahoo.com).