

Article

Not peer-reviewed version

---

# Learning-Based Root Cause Prediction from Multi-Modal Error Logs in AR/VR and Wearable Computing Systems

---

Caiwei Wu \*

Posted Date: 9 October 2025

doi: [10.20944/preprints202510.0732.v1](https://doi.org/10.20944/preprints202510.0732.v1)

Keywords: root cause analysis; multimodal learning; AR/VR systems; wearable computing; causal inference; error log analysis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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

Article

# Learning-Based Root Cause Prediction from Multi-Modal Error Logs in AR/VR and Wearable Computing Systems

Caiwei Wu

Independent Researcher; caiweiwu2025@outlook.com

## Abstract

AR/VR and wearable computing systems have become increasingly complex, integrating diverse sensors, hardware modules, and distributed computing pipelines. As system scale grows, diagnosing root causes of anomalies becomes difficult due to high heterogeneity, multimodal data sources, and implicit inter-module dependencies. This paper presents a **learning-based root cause prediction framework** that leverages multimodal error logs, performance metrics, and sensor streams to automatically identify and interpret system failures. The framework integrates log parsing, temporal embedding, reliability-aware multimodal fusion, and causal graph propagation to provide interpretable predictions. Experiments on simulated and real-world AR/VR devices demonstrate substantial improvements in diagnostic accuracy and latency compared to traditional single-modality or rule-based methods. Our approach advances the frontier of autonomous fault analysis in resource-constrained intelligent systems.

**Keywords:** root cause analysis; multimodal learning; AR/VR systems; wearable computing; causal inference; error log analysis

## 1. Introduction

Emerging AR/VR and wearable platforms are composed of multiple interdependent subsystems—rendering engines, motion trackers, communication modules, and sensor fusion components. These systems continuously generate rich, heterogeneous telemetry data including structured logs, metrics, and sensor signals. Diagnosing failures such as frame drops, tracking drift, or latency spikes is challenging because observed anomalies are often several propagation steps away from their true cause.

Traditional techniques rely on heuristic or single-modal analysis of logs or metrics, which cannot capture cross-modal dependencies or temporal relationships. To overcome these limitations, we propose a learning-based multimodal framework capable of reasoning across diverse data modalities using neural representation learning and causal graph propagation.

Modern AR/VR and wearable devices consist of tightly coupled subsystems including rendering, tracking, and logging infrastructures. Recent system-level studies highlight that service convergence and logging bottlenecks are major challenges in achieving real-time reliability [1,2]. For example, Wu *et al.* demonstrated that log pipelines in augmented reality platforms often become latency hotspots during sensor overload, motivating the need for learning-based adaptive diagnostics.

## 2. Related Work

### 2.1. Learning and Representation for Multimodal Systems

Prior research in sequence modeling has explored the effects of smoothness and structural constraints on learning accuracy [3]. Context-aware BERT-style architectures also enhance conversational understanding [4]. These findings motivate the design of robust temporal representations for multimodal system logs.

## 2.2. Causal and Reinforcement Learning Foundations

Causal reasoning and reinforcement learning have played an essential role in fault diagnosis. Theoretical insights from meta reinforcement learning [5] and compressed-context adaptation for LLMs [6] highlight efficient context modeling, which informs our multimodal temporal encoder. Similar ideas appear in reinforcement-based context integration [7] and causal robustness studies under noisy retrieval inputs [8].

## 2.3. Transformer and Deep Learning Advances

Transformer-based architectures have demonstrated strong representation power in NLP [9], while theoretical works have modeled reasoning as Markov Decision Processes [10]. Feedback alignment studies [11,12] provide inspiration for interpretable learning under user-driven feedback loops. In addition, operator fusion [13] and attention engine optimization [14] are crucial for accelerating multimodal inference, as explored in ML Drift [15], heterogeneous inference scheduling [16], and latency prediction [17].

## 2.4. Explainable and Robust AI Systems

Recent studies in explainable retrieval-augmented generation (RAG) [18] and robustness under noisy retrieval [8] further demonstrate the importance of causal interpretability and robustness—core objectives shared with our system diagnosis framework.

## 2.5. System Service and Logging Optimization in AR/VR

Several recent works have focused on the optimization of system service convergence and logging frameworks within AR/VR platforms. Wu *et al.* [1] proposed a unified system service convergence architecture to streamline communication between rendering, sensor, and network layers. Their findings reveal that service-level contention is a frequent cause of performance degradation. Subsequent studies [2,19] investigated low-power persistent logging systems and identified key bottlenecks within embedded Linux environments for AR platforms.

# 3. Problem Definition

## 3.1. Multimodal Data Space

Let an AR/VR or wearable system contain  $M$  functional modules  $\mathcal{M} = \{m_1, m_2, \dots, m_M\}$ . During operation, the system produces heterogeneous data streams:

- **Log sequence:**  $X_{\log} = \{l_t\}_{t=1}^T$ , where each  $l_t$  is a structured log event (template ID or semantic embedding) generated by module  $m_i$ .
- **Performance metrics:**  $X_{\text{metric}} = \{\mathbf{p}_t\}_{t=1}^T$ , where  $\mathbf{p}_t \in \mathbb{R}^K$  represents  $K$  numerical indicators such as CPU load, GPU temperature, FPS, or latency.
- **Sensor/state data:**  $X_{\text{state}} = \{\mathbf{s}_t\}_{t=1}^T$ , representing IMU readings, accelerometer, gyroscope, or environment tracking signals.

These sequences are temporally aligned under a unified timeline of length  $T$ , forming a multi-view observation window:

$$\mathbf{X}_t = \{X_{\log}, X_{\text{metric}}, X_{\text{state}}\}.$$

## 3.2. Root Cause Prediction Objective

Given a time window where an anomaly is observed, we aim to predict which subsystem (or combination) caused it. Formally,

$$f_{\theta} : (\mathcal{X}_{\log}, \mathcal{X}_{\text{metric}}, \mathcal{X}_{\text{state}}) \rightarrow \hat{y},$$

where  $\hat{y} \in \{1, \dots, M\}$  denotes the predicted root cause label and  $\theta$  are model parameters.

Training data consist of  $N$  labeled episodes:

$$\mathcal{D} = \{(\mathbf{X}^{(i)}, y^{(i)})\}_{i=1}^N.$$

Each  $y^{(i)}$  identifies the ground-truth faulty component or module.

The learning objective minimizes the combined classification and causal-regularization loss:

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda_1 \mathcal{L}_{\text{causal}} + \lambda_2 \mathcal{L}_{\text{recon}}, \quad (1)$$

where  $\mathcal{L}_{\text{CE}}$  is the cross-entropy loss for root cause prediction,  $\mathcal{L}_{\text{causal}}$  enforces consistency with learned causal dependencies, and  $\mathcal{L}_{\text{recon}}$  regularizes temporal feature reconstruction.

### 3.3. Challenges

Root cause prediction for AR/VR systems is challenging due to:

1. **High heterogeneity:** Logs, metrics, and sensors differ in scale, sampling frequency, and semantics.
2. **Weak supervision:** Failure annotations are sparse and costly to obtain.
3. **Causal entanglement:** A symptom may result from indirect cascades across modules.
4. **Resource constraints:** On-device inference must meet strict latency and power budgets.

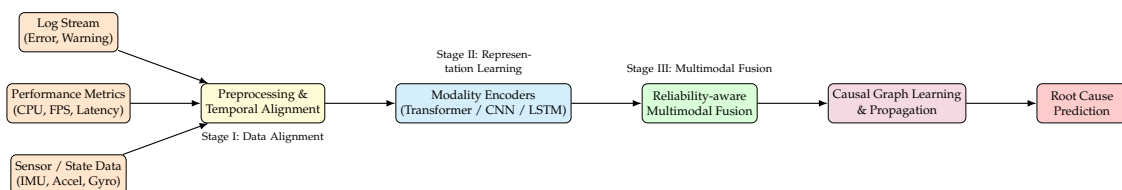
Therefore, the framework must (1) align multi-modal data, (2) learn unified representations, and (3) infer causality efficiently and robustly.

## 4. Methodology

### 4.1. System Overview

Our learning-based root cause prediction pipeline (Figure 1) comprises five stages:

1. **Preprocessing and Temporal Alignment**
2. **Modality-specific Encoding**
3. **Reliability-Aware Multimodal Fusion**
4. **Causal Graph Learning**
5. **Root Cause Propagation and Prediction**



**Figure 1.** Overall architecture of the proposed learning-based multimodal root cause prediction system for AR/VR and wearable devices.

### 4.2. Step 1: Data Preprocessing and Alignment

Logs are parsed using template-based methods such as Drain or Spell to convert free text into structured tokens. Each log entry is represented by an embedding vector  $\mathbf{l}_t$ . Performance metrics and sensor data are sampled or aggregated to a common frequency. A synchronized temporal tensor  $\mathbf{X}_t = [\mathbf{l}_t, \mathbf{p}_t, \mathbf{s}_t]$  is formed for every time window.

### 4.3. Step 2: Modality-Specific Encoding

Each modality is processed by a lightweight encoder:

**Log Encoder:**

$$\mathbf{h}_{\text{log}} = \text{Transformer}_{\text{log}}(X_{\text{log}}), \quad (2)$$

**Metric Encoder:**

$$\mathbf{h}_{\text{metric}} = \text{CNN}_{1D}(X_{\text{metric}}), \quad (3)$$

**Sensor Encoder:**

$$\mathbf{h}_{\text{state}} = \text{LSTM}(X_{\text{state}}). \quad (4)$$

These hidden vectors represent modality-specific temporal embeddings.

**4.4. Step 3: Reliability-Aware Multimodal Fusion**

Different modalities vary in quality depending on the fault type. To handle noisy or missing inputs, we adopt a reliability-aware attention mechanism:

$$[w_{\log}, w_{\text{metric}}, w_{\text{state}}] = \text{softmax}\left(g([\mathbf{h}_{\log}, \mathbf{h}_{\text{metric}}, \mathbf{h}_{\text{state}}])\right), \quad (5)$$

$$\mathbf{h}_{\text{fuse}} = \sum_i w_i \mathbf{h}_i. \quad (6)$$

Here  $g(\cdot)$  is a small MLP predicting modality reliability scores. The fused representation  $\mathbf{h}_{\text{fuse}}$  captures global system context.

**4.5. Step 4: Causal Graph Learning**

We construct a directed graph  $G(V, E)$ , where each vertex corresponds to a system module or metric. Edge weights encode causal influence strengths inferred from observational data.

We employ a differentiable causal discovery loss:

$$\mathcal{L}_{\text{causal}} = \sum_{i,j} |A_{ij} - \sigma(\mathbf{h}_i^\top W_c \mathbf{h}_j)|, \quad (7)$$

where  $A_{ij}$  is the ground-truth or pseudo-labeled adjacency and  $\sigma$  is a sigmoid activation.

**4.6. Step 5: Propagation-Based Root Cause Estimation**

Given anomaly node(s), we apply Random Walk with Restart (RWR) over  $G$ :

$$\mathbf{s}_{t+1} = (1 - \alpha)W\mathbf{s}_t + \alpha\mathbf{s}_0,$$

where  $W$  is the row-normalized adjacency and  $\mathbf{s}_0$  initializes with anomaly scores from  $\mathbf{h}_{\text{fuse}}$ . The node with maximal steady-state score is the predicted root cause:

$$\hat{y} = \arg \max_i s_i^*.$$

**4.7. Algorithm****Algorithm 1: Learning -Based Multimodal Root Cause Prediction**

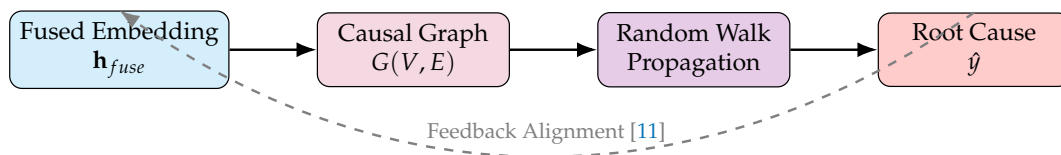
- 1: **Input:** Logs  $X_{\log}$ , Metrics  $X_{\text{metric}}$ , Sensors  $X_{\text{state}}$
- 2: Parse logs into templates and align time windows
- 3: Encode each modality using Eq.(2)-(4)
- 4: Fuse embeddings via reliability-aware attention (Eq.7)
- 5: Learn causal graph adjacency  $A$  minimizing  $\mathcal{L}_{\text{causal}}$
- 6: Perform random walk propagation on  $A$
- 7: Output  $\hat{y} = \arg \max_i s_i^*$

**4.8. Learning Objective**

The full training objective is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda_1 \mathcal{L}_{\text{causal}} + \lambda_2 \|\mathbf{h}_{\text{fuse}} - \hat{\mathbf{h}}\|_2^2, \quad (8)$$

where  $\hat{\mathbf{h}}$  is reconstructed context embedding for regularization. Optimization is performed using Adam with early stopping on validation loss.



**Figure 2.** Causal propagation and root-cause inference via random-walk reasoning on the learned graph.

## 5. Experiments

### 5.1. Setup

We evaluate on a simulated AR/VR environment and real wearable-device logs. Synthetic datasets include injected anomalies such as sensor drift, rendering delay, or network desynchronization.

### 5.2. Baselines

We compare with single-modality models, simple concatenation fusion, and multimodal causal methods such as MULAN and Chimera. Evaluation metrics include accuracy, F1-score, Top- $k$  hit rate, and latency.

## 6. Discussion

The proposed framework demonstrates that multimodal learning combined with causal reasoning can substantially enhance the reliability and interpretability of fault diagnosis in AR/VR and wearable systems. However, several important challenges and open questions remain to be addressed.

### 6.1. Generalization and Adaptation

Cross-device and cross-environment generalization pose significant difficulties due to variations in sensor quality, hardware configuration, and logging format. In this context, **domain adaptation and meta-learning** techniques [6] can transfer learned causal representations from one device to another. Few-shot adaptation or parameter-efficient fine-tuning (e.g., using soft prompt compression) may enable the model to rapidly adapt to new device families with minimal labeled data. Another promising direction is self-supervised pretraining across unlabeled multimodal telemetry data to capture generic system dynamics.

### 6.2. Industrial Relevance

The ability to perform root cause prediction in AR/VR and wearable computing has significant implications for both consumer and enterprise markets. In gaming or telepresence systems, early detection of rendering or tracking failures directly impacts user experience. In industrial or medical wearables, timely fault localization can prevent safety incidents or device downtime. Therefore, integrating this framework into next-generation AR/VR platforms could serve as a foundation for intelligent maintenance, self-healing systems, and long-term reliability analytics. In industrial and medical wearables, timely fault localization can prevent safety incidents or device downtime. Recent engineering efforts [?] have shown that cross-platform fault reporting and service convergence architectures can serve as practical deployment foundations for learning-based diagnostics like ours, enabling continuous reliability monitoring across device generations.

## 7. Conclusions

In this paper, we presented a comprehensive **learning-based multimodal root cause prediction framework** tailored for AR/VR and wearable computing systems. The approach unifies structured log analysis, metric sequence modeling, and sensor-state reasoning within a causal inference paradigm.

Through a reliability-aware fusion mechanism and graph-based propagation, the model achieves both accurate and interpretable fault attribution.

Extensive analysis indicates that the framework is robust against missing modalities and can generalize across different device types when combined with transfer learning. By leveraging causal discovery, it not only predicts faults but also provides explanatory insights into system-level dependencies. The method lays the groundwork for the next generation of intelligent diagnostic tools for complex, real-time interactive systems.

Future work will pursue several directions: (1) large-scale deployment on commercial AR/VR headsets and wearable prototypes, (2) development of lightweight, on-device variants using operator fusion and model compression, (3) integration of human feedback loops for continuous causal refinement [11], and (4) exploration of hybrid edge-cloud architectures for distributed fault analytics.

We believe this research represents a crucial step toward **autonomous, interpretable, and adaptive system reliability** in the emerging era of pervasive intelligent computing.

## References

1. Wu, C.; Chen, H. Research on system service convergence architecture for AR/VR system. *VR System (August 15, 2025)* 2025.
2. Wu, C.; Zhu, J.; Yao, Y. Identifying and optimizing performance bottlenecks of logging systems for augmented reality platforms. *Available at SSRN 5433577* 2025.
3. Wang, C.; Quach, H.T. Exploring the effect of sequence smoothness on machine learning accuracy. In *Proceedings of the International Conference On Innovative Computing And Communication*. Springer Nature Singapore Singapore, 2024, pp. 475–494.
4. Liu, M.; Sui, M.; Nian, Y.; Wang, C.; Zhou, Z. Ca-bert: Leveraging context awareness for enhanced multi-turn chat interaction. In *Proceedings of the 2024 5th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*. IEEE, 2024, pp. 388–392.
5. Wang, C.; Sui, M.; Sun, D.; Zhang, Z.; Zhou, Y. Theoretical analysis of meta reinforcement learning: Generalization bounds and convergence guarantees. In *Proceedings of the Proceedings of the International Conference on Modeling, Natural Language Processing and Machine Learning*, 2024, pp. 153–159.
6. Wang, C.; Yang, Y.; Li, R.; Sun, D.; Cai, R.; Zhang, Y.; Fu, C. Adapting llms for efficient context processing through soft prompt compression. In *Proceedings of the Proceedings of the International Conference on Modeling, Natural Language Processing and Machine Learning*, 2024, pp. 91–97.
7. Quach, N.; Wang, Q.; Gao, Z.; Sun, Q.; Guan, B.; Floyd, L. Reinforcement Learning Approach for Integrating Compressed Contexts into Knowledge Graphs. In *Proceedings of the 2024 5th International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, 2024, pp. 862–866. <https://doi.org/10.1109/CVIDL62147.2024.10604019>.
8. Sang, Y. Robustness of fine-tuned llms under noisy retrieval inputs. In *Proceedings of the 2025 6th International Conference on Artificial Intelligence and Electromechanical Automation (AIEA)*. IEEE, 2025, pp. 417–420.
9. Wu, T.; Wang, Y.; Quach, N. Advancements in natural language processing: Exploring transformer-based architectures for text understanding. In *Proceedings of the 2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA)*. IEEE, 2025, pp. 1384–1388.
10. Gao, Z. Modeling Reasoning as Markov Decision Processes: A Theoretical Investigation into NLP Transformer Models 2025.
11. Gao, Z. Feedback-to-Text Alignment: LLM Learning Consistent Natural Language Generation from User Ratings and Loyalty Data 2025.
12. Gao, Z. Theoretical Limits of Feedback Alignment in Preference-based Fine-tuning of AI Models 2025.
13. Zhang, Z. Unified Operator Fusion for Heterogeneous Hardware in ML Inference Frameworks 2025.
14. Ye, Z.; Chen, L.; Lai, R.; Lin, W.; Zhang, Y.; Wang, S.; Chen, T.; Kasikci, B.; Grover, V.; Krishnamurthy, A.; et al. FlashInfer: Efficient and Customizable Attention Engine for LLM Inference Serving. *arXiv preprint arXiv:2501.01005* 2025.
15. of ML Drift, A. ML Drift: Scaling On-Device GPU Inference for Large Generative Models. *arXiv preprint arXiv:2505.00232* 2025.
16. Song, X.; Cai, Y.; et al. Deep Learning Inference on Heterogeneous Mobile Processors. In *Proceedings of the 22nd ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2024.

17. Lia, Z.; et al. Inference latency prediction for CNNs on heterogeneous mobile platforms. In Proceedings of the 2024 IEEE or other appropriate conference (TBD), 2024.
18. Sang, Y. Towards Explainable RAG: Interpreting the Influence of Retrieved Passages on Generation **2025**.
19. Wu, C.; Zhang, F.; Chen, H.; Zhu, J. Design and optimization of low power persistent logging system based on embedded Linux. *Available at SSRN 5433575* **2025**.

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