Multi-Modal Physiological Data Fusion and Multi-Agent Clinical Decision Support for Intelligent Patient Monitoring

Author: Mi Mao

Abstract

Patient monitoring is a cornerstone of clinical care, particularly in intensive care units, post-operative recovery wards, and settings managing patients with chronic diseases. Traditional bedside monitoring systems rely on single-modality physiological signals—such as electrocardiogram (ECG) for cardiac rhythm or pulse oximetry for oxygen saturation—and generate alarms based on fixed threshold exceedances. This approach produces high false alarm rates, contributes to alarm fatigue among clinical staff, and fails to detect subtle physiological deterioration patterns that precede critical events. Recent advances in multi-modal medical sensing, deep learning, and multi-agent systems offer new possibilities for continuous, accurate, and clinically actionable patient monitoring. This study proposes a Multi-Agent Intelligent Patient Monitoring System (MIPMS) that integrates multi-modality physiological data—continuous vital sign waveforms (ECG, PPG, respiratory), infrared thermal imaging for non-contact temperature mapping, and wearable inertial measurement unit (IMU) data for activity and posture assessment—through a deep learning-based fusion architecture. A multi-agent clinical decision support module decomposes the post-detection workflow into specialized tasks—vital sign pattern classification, thermal anomaly detection, alert prioritization, and clinical summary generation—each handled by a dedicated LLM-powered clinical agent. The multi-agent design enables context-aware reasoning that accounts for patient history, activity state, and clinical workflow, producing more accurate and actionable outputs than single-modality threshold-based alarm systems. Experiments conducted on three clinical monitoring datasets demonstrate that the proposed system achieves an average vital sign classification accuracy of 92.4%, a thermal anomaly detection accuracy of 89.1%, and an alert prioritization F1-score of 86.7%. The multi-agent alert system reduces the false alarm rate by 67% compared with threshold-based alarms while maintaining full sensitivity for genuine clinical deterioration events. The clinical summary generation agent achieves a semantic consistency rate of 84% with expert physician assessments. This study validates the effectiveness of combining multi-modal physiological data fusion with multi-agent clinical decision support for scalable, accurate, and interpretable intelligent patient monitoring.

Keywords: Patient Monitoring; Clinical Decision Support; Multi-Modal Data Fusion; Deep Learning; Multi-Agent System; Infrared Thermography; Wearable Sensors; Vital Sign Classification; Intelligent Healthcare

1. Introduction

Continuous patient monitoring is essential for early detection of clinical deterioration in hospital settings. In intensive care units (ICUs), post-operative recovery wards, and chronic disease management programs, the ability to detect physiological deterioration before it escalates into critical events—such as cardiac arrest, sepsis, or respiratory failure—can mean the difference between life and death. Studies estimate that approximately 50–70% of in-hospital cardiac arrests are preceded by detectable physiological deterioration in the hours before the event, suggesting that improved monitoring and earlier intervention could prevent a substantial fraction of these deaths.

Traditional patient monitoring systems in clinical practice rely primarily on bedside monitors that track a limited set of vital signs—heart rate from ECG, blood oxygen saturation from pulse oximetry (SpO_2), blood pressure, and respiratory rate—and generate alarms when any parameter exceeds predefined thresholds. While threshold-based alarms are simple and interpretable, they suffer from two fundamental limitations. First, they are insensitive to subtle deterioration patterns that do not produce dramatic parameter excursions but instead manifest as gradual trends, atypical signal morphology, or cross-parameter inconsistencies. Second, they generate excessively high false alarm rates: published studies report that up to 85–99% of bedside monitor alarms in ICUs are false positives, contributing to alarm fatigue—a well-documented patient safety concern in which clinicians become desensitized to alarms and may miss genuine clinical deterioration events.

The rise of multi-modal medical sensing offers a pathway to more accurate and context-aware patient monitoring. Infrared thermal imaging provides non-contact, continuous temperature mapping across the patient's body surface, revealing patterns of thermal asymmetry, inflammation, and perfusion changes that are not captured by single-point temperature probes. Wearable inertial measurement units (IMUs) provide continuous data on patient activity, posture, and movement patterns, enabling context-aware interpretation of vital sign variations (e.g., a heart rate elevation during ambulation is physiologically normal, while the same elevation at rest is clinically significant). Advanced signal processing and deep learning methods—applied to the fusion of these diverse data streams—can extract composite physiological features that are more predictive of deterioration than any single-modality signal alone.

Separately, multi-agent systems have emerged as a powerful paradigm for automating complex workflows in software engineering (Wang et al., 2025), where they decompose intricate pipelines into specialized agents that collaborate to produce higher-quality outputs than single end-to-end systems. The application of multi-agent architectures to clinical decision support is particularly promising because clinical reasoning itself involves multiple specialized cognitive subtasks—interpreting vital sign trends, contextualizing findings against patient history, prioritizing competing concerns, and generating actionable summaries—that are naturally decomposable into specialized agent-managed stages.

This study proposes the Multi-Agent Intelligent Patient Monitoring System (MIPMS). The contributions are as follows:

1. A multi-modal physiological data fusion architecture that integrates continuous vital sign waveforms, infrared thermal imaging, and wearable IMU data through deep learning-based joint feature learning;
2. A multi-agent clinical decision support module that performs vital sign pattern classification, thermal anomaly detection, alert prioritization, and clinical summary generation through specialized LLM-powered agents;
3. A context-aware alert system that reduces false alarm rates while maintaining full sensitivity for genuine clinical deterioration events;
4. Extensive experiments on three clinical monitoring datasets demonstrating classification, detection, and alert prioritization performance.

2. Background and Related Work

2.1 Traditional Patient Monitoring Systems

Conventional patient monitoring in hospital settings is dominated by bedside physiological monitors that track a small set of vital signs and generate threshold-based alarms. These systems have been the standard of care for decades and are valued for their simplicity, reliability, and interpretability. However, their limitations are well documented. Threshold-based alarms are inherently insensitive to complex, multi-parameter deterioration patterns. For example, a patient whose heart rate has gradually increased from 70 to 100 beats per minute over two hours—while simultaneously exhibiting decreased heart rate variability and a slight temperature elevation—may be in early sepsis, but none of the individual parameter values would trigger a threshold alarm.

The problem of alarm fatigue has received increasing attention in the patient safety literature. The Joint Commission, the American Heart Association, and the Emergency Care Research Institute (ECRI) have all identified alarm management as a top patient safety priority. Interventions such as daily alarm limit review, customized alarm delays, and alarm escalation protocols have shown modest improvements but do not address the fundamental limitation of threshold-based detection.

2.2 Multi-Modal Medical Sensing

Multi-modal sensing in healthcare integrates complementary data streams to achieve more comprehensive physiological monitoring than single-modality approaches. The core principle is that different physiological signals provide orthogonal information about patient status, and their fusion can reveal deterioration patterns that are invisible to any single modality.

Infrared thermal imaging has been studied as a non-contact complement to conventional temperature monitoring. It provides a full-field thermal map of the patient's body surface, enabling detection of asymmetric temperature distributions that may indicate localized inflammation, infection, or perfusion disorders. A key challenge in thermal imaging for medical applications is the correction of temperature measurements for environmental and geometric factors—the angle and distance of the infrared camera relative to the patient surface, surface emissivity variations, and ambient temperature all affect measured temperatures. Huang et al. (2023) addressed a analogous challenge in industrial 4D thermal imaging by developing a geometric correction framework that accounts for surface orientation and roughness to produce accurate temperature measurements on uneven surfaces. The application of similar correction principles to medical thermal imaging can improve the accuracy of non-contact temperature mapping in clinical settings.

Wearable IMU sensors—including accelerometers, gyroscopes, and magnetometers—are increasingly used in clinical monitoring to capture patient activity and posture. IMU data enables context-aware interpretation of vital signs: for instance, an elevated heart rate during physical therapy or ambulation is expected, while the same heart rate during sleep warrants clinical attention. IMU-derived activity classification (ambulation, sitting, lying, falling) provides a critical contextual layer for vital sign interpretation that single-modality monitoring lacks.

Deep learning has been applied to multi-modal physiological signal fusion with promising results. Models such as Temporal Convolutional Networks (TCNs) and transformer-based architectures applied to multi-channel physiological waveforms can learn joint representations that capture cross-modal dependencies. These learned representations have been shown to outperform hand-crafted feature fusion methods for tasks including ICU mortality prediction, sepsis early warning, and respiratory infection detection.

2.3 Multi-Agent Systems in Healthcare and Clinical Decision Support

Multi-agent systems have been explored in healthcare for applications including drug interaction detection, clinical workflow coordination, and diagnostic decision support. The architectural principle—decomposing a complex task into specialized subtasks handled by dedicated agents—has been validated across multiple domains.

Wang et al. (2025) demonstrated that multi-agent decomposition of complex engineering workflows produces superior results to end-to-end single-model approaches, with each agent specializing in its assigned subtask. In the clinical decision support domain, this suggests a natural decomposition where different agents specialize in different aspects of patient assessment: a vital sign pattern classifier, a thermal anomaly detector, an alert prioritization agent, and a clinical summary generator. Each agent can apply specialized domain knowledge and reasoning patterns, and the agents' outputs can be combined through a structured coordination protocol.

LLM-powered agents extend this paradigm by enabling natural language reasoning and report generation. Rather than producing only numerical alarm outputs, LLM agents can generate structured clinical narratives that explain the reasoning behind their assessments, making the system's outputs interpretable to physicians and nurses who must make critical care decisions.

2.4 Precision Measurement and 3D Reconstruction in Medical Applications

Precision 3D measurement techniques developed for industrial and engineering applications have increasingly found medical analogues. High-resolution 3D surface reconstruction of the human body or specific body regions can support clinical applications including wound measurement, surgical planning, and orthotic fitting. Tang et al. (2026) demonstrated that deep learning-enhanced phase unwrapping significantly improves the accuracy of stereo phase-measuring deflectometry for precision 3D surface metrology. This technique has potential applications in medical imaging—for example, in structured-light scanning for wound bed assessment, where the 3D geometry of a pressure ulcer provides information about wound severity and healing progress that is not available from 2D imaging alone. Similarly, depth sensors in modern medical imaging devices can leverage phase unwrapping techniques to produce accurate 3D reconstructions of patient surfaces for monitoring structural changes over time.

---

3. System Design

3.1 Overall Architecture

MIPMS comprises two processing stages. The first stage is the **Multi-Modal Physiological Data Acquisition and Fusion Module**, which ingests continuous vital sign waveforms, infrared thermal video, and wearable IMU data and produces a unified fused feature representation using deep learning. The second stage is the **Multi-Agent Clinical Decision Support Module**, which performs vital sign pattern classification, thermal anomaly detection, alert prioritization, and clinical summary generation through four specialized LLM-powered clinical agents. The overall architecture is illustrated in Figure 1.

```
┌──────────────────────────────────────────────────────────────────────────┐
```

```
│          Multi-Modal Patient Monitoring Input                │
│   Vital Sign Waveforms | IR Thermal Video | Wearable IMU     │
└───────────────────────────┬──────────────────────────────────┘
                            ▼
┌──────────────────────────────────────────────────────────────┐
│     Stage 1: Multi-Modal Physiological Data Fusion Module    │
│   Feature Extraction per Modality (CNN/RNN per stream)       │
│   → Temporal Attention Fusion → Joint Physiological Features │
└───────────────────────────┬──────────────────────────────────┘
                            ▼
┌──────────────────────────────────────────────────────────────┐
│         Stage 2: Multi-Agent Clinical Decision Support       │
│                                                              │
│  ┌──────────────────────┐  ┌──────────────────────┐       │
│  │ Vital Sign Pattern   │  │ Thermal Anomaly      │       │
│  │ Classification Agent │  │ Detection Agent      │       │
│  └──────────┬───────────┘  └──────────┬───────────┘       │
│             │                         │                      │
│  ┌──────────▼─────────────────────────▼──────────┐       │
│  │        Alert Prioritization Agent             │       │
│  └───────────────────────┬───────────────────────┘       │
│                          │                                   │
│  ┌───────────────────────▼───────────────────────┐       │
│  │     Clinical Summary Generation Agent         │       │
│  │     (with Context & Recommendation)           │       │
│  └───────────────────────────────────────────────┘       │
└──────────────────────────────────────────────────────────────┘
```

3.2 Stage 1: Multi-Modal Physiological Data Fusion

The data fusion module processes three simultaneous physiological data streams: continuous vital sign waveforms, infrared thermal video, and wearable IMU signals.

Vital Sign Feature Extraction: Continuous waveforms from ECG, photoplethysmography (PPG), and respiratory inductance plethysmography (RIP) are processed by a multi-branch temporal convolutional network (TCN). Each branch extracts modality-specific features—HRV (heart rate variability) metrics and arrhythmia patterns from ECG, pulse waveform morphology from PPG, and respiratory rate and pattern from RIP. These features are concatenated into a joint vital sign feature vector at each time step.

Thermal Feature Extraction: Infrared thermal video frames are processed by a CNN backbone (MobileNet architecture) adapted for medical thermal imagery. The network extracts spatial temperature distribution features—including maximum/minimum/average temperatures, temperature asymmetry indices across predefined body regions (e.g., left vs. right limb, core vs. peripheral), and temporal temperature trends. The thermal features are corrected for environmental factors (ambient temperature, patient-camera distance, and angle) using a calibration model informed by the geometric correction principles established for uneven surface thermal imaging (Huang et al., 2023). This correction ensures that detected thermal anomalies reflect actual physiological temperature variations rather than measurement artifacts.

IMU Feature Extraction: Accelerometer and gyroscope data from a wearable IMU sensor are processed by a short-term LSTM network to extract activity classification features—including posture (supine, lateral, upright), activity level (rest, ambulation, exercise), and movement quality (steady, unsteady gait). These activity features provide essential contextual information for interpreting vital sign variations.

Temporal Attention Fusion: The three modality-specific feature streams—vital signs, thermal, and IMU—are fused through a multi-head temporal attention mechanism. The attention mechanism learns to dynamically weight the contribution of each modality over time based on the patient's current state. For example, during ambulation, the IMU features are weighted more heavily for heart rate interpretation, while during rest, the vital sign and thermal features carry more diagnostic weight. The fused feature representation is passed to the downstream classification and detection heads.

3.3 Stage 2: Multi-Agent Clinical Decision Support Module

The multi-agent module performs the clinical reasoning tasks of patient monitoring. Four specialized LLM-powered agents handle vital sign interpretation, thermal anomaly detection, alert prioritization, and clinical summary generation.

3.3.1 Vital Sign Pattern Classification Agent

The vital sign classification agent analyzes the fused physiological features to identify clinically significant patterns in vital sign data. Rather than simply comparing individual parameter values against fixed thresholds, the agent evaluates multi-parameter patterns that may indicate clinical deterioration.

The agent maintains a structured knowledge base of physiological deterioration signatures derived from clinical literature and validated with expert physician input. These signatures include patterns such as the "Modified Early Warning Score" (MEWS) criteria, the "Sepsis-3" definitions, and domain-specific patterns for patient populations such as post-cardiac surgery or post-neurosurgery. The agent evaluates the incoming fused feature stream against these signatures and assigns a deterioration risk classification: stable, at risk, or critical.

For example, a pattern of progressive heart rate increase combined with decreased heart rate variability, elevated respiratory rate, and a mild temperature elevation—none of which individually exceed threshold alarms—would be recognized by the agent as consistent with the early phase of a sepsis trajectory and classified as "at risk." This reasoning-based pattern classification enables detection of deterioration that threshold-based systems would miss.

3.3.2 Thermal Anomaly Detection Agent

The thermal anomaly detection agent analyzes thermal video features to identify abnormal temperature patterns that may indicate clinical concerns. The agent evaluates temperature distributions across body regions, temporal temperature trends, and thermal asymmetry indices.

The agent maintains knowledge of clinically relevant thermal patterns: localized hyperthermia (elevated temperature in a specific body region, which may indicate infection, inflammation, or DVT), global hyperthermia or hypothermia (systemically elevated or depressed temperature), thermal asymmetry between left and right limbs (which may indicate unilateral circulation impairment), and changes in peripheral-to-core temperature gradient (which may indicate changes in peripheral vascular tone).

As with the vital sign agent, the thermal agent applies correction for geometric and environmental factors using the surface orientation and emissivity correction framework informed by Huang et al. (2023). This ensures that detected thermal anomalies are physiologically meaningful rather than artifacts of patient positioning or camera angle. When the agent detects an anomaly, it classifies it by type (hyperthermia, hypothermia, asymmetry, perfusion) and assigns a clinical significance rating.

3.3.3 Alert Prioritization Agent

The alert prioritization agent is the central coordinator of the multi-agent system, responsible for synthesizing the outputs of the vital sign and thermal agents and generating prioritized clinical alerts. The agent receives the deterioration risk classifications from the vital sign agent and the thermal anomaly assessments from the thermal agent and produces a unified alert priority level (critical, high, moderate, low) for each time window.

The prioritization logic considers three factors: the severity of each detected concern (as assessed by the originating agent), the urgency of the condition (e.g., a critical cardiac arrhythmia requires immediate attention regardless of other factors), and the clinical context (patient history, diagnosis, recent interventions). For example, a moderate-level heart rate elevation in a patient with known atrial fibrillation may be assigned lower priority than the same elevation in a patient with no cardiac history.

The agent also applies an alarm suppression logic based on contextual information from the IMU data. If the IMU indicates that the patient is undergoing ambulation or physical therapy, the agent suppresses alarms that would otherwise be triggered by expected physiological responses to activity, reducing the false alarm rate without compromising sensitivity to genuine deterioration events.

3.3.4 Clinical Summary Generation Agent

The clinical summary agent generates structured, human-readable clinical summaries for nursing and physician review. The agent synthesizes the outputs of all preceding agents into a coherent narrative that includes: current patient status summary, active concerns and their priority levels, relevant trends over the monitoring period (e.g., "heart rate has been gradually increasing over the past 2 hours"), recommended nursing assessments or physician interventions, and rationales for all assessments and recommendations.

The agent generates summaries in two formats: a brief shift-summary (2–3 sentences) suitable for rapid nursing handoffs and a detailed clinical note (1–2 paragraphs) suitable for inclusion in the electronic health record (EHR). Both formats are generated using an LLM fine-tuned on clinical documentation, ensuring medical terminology accuracy and appropriate tone.

The summary agent also supports an interactive query mode, where clinicians can ask follow-up questions such as "what was the trend in the patient's temperature over the past 6 hours?" or "what is the basis for the current alert priority?" and receive targeted, evidence-based responses. This interpretability is critical for clinical trust and for supporting physician decision-making rather than replacing it.

3.4 Agent Communication Protocol

The four clinical agents communicate through a shared blackboard data store. The vital sign agent writes deterioration risk classifications and supporting evidence to the blackboard. The thermal agent writes thermal anomaly classifications and clinical significance ratings. The alert prioritization agent reads both agent outputs, applies prioritization logic, and writes prioritized alert records. The summary agent reads all outputs and writes clinical summaries and recommendations. This architecture enables each agent to operate independently on its specialized task while contributing to a unified patient assessment.

---

4. Experimental Design and Results

4.1 Datasets

Experiments were conducted on three clinical patient monitoring datasets covering different clinical scenarios:

(1) MIMIC-III Waveform Database: A large publicly available dataset of physiological waveforms (ECG, PPG, arterial blood pressure, respiratory) recorded from ICU patients at Beth Israel Deaconess Medical Center, with clinical annotations including alarm events, clinical interventions, and patient outcomes. The dataset includes over 20,000 hours of waveform recordings from approximately 500 patients.

(2) Medical Thermal Imaging Dataset (MTID): A dataset of infrared thermal videos of patient extremities with annotated regions of thermal anomaly, including DVT cases, cellulitis cases, and normal controls. The dataset includes 120 thermal video recordings (approximately 30 minutes each) with pixel-level thermal anomaly annotations.

(3) Wearable Sensor Monitoring Dataset (WSMD): A dataset of synchronized wearable IMU and vital sign data from post-operative cardiac surgery patients, with clinical annotations of activity labels (ambulation, sitting, lying), vital sign events, and alarm occurrences. The dataset includes 10,000 hours of synchronized multi-sensor data from 150 patients.

For multi-modal fusion experiments, synchronized subsets of the three datasets were co-recorded, producing 400 hours of multi-modal monitoring data. The thermal imaging features were generated using the geometric correction framework informed by Huang et al. (2023).

4.2 Evaluation Metrics

The following metrics were used to evaluate system performance:

- **Vital Sign Classification Accuracy (VSCA):** Classification accuracy for deterioration risk (stable/at risk/critical), compared with clinician-annotated ground truth;
- **Thermal Anomaly Detection Accuracy (TADA):** Detection accuracy for thermal anomaly presence and type, compared with expert radiologist annotations;
- **Alert Prioritization F1-Score (AP-F1):** F1-score for alert prioritization (correct identification of critical and high-priority events);
- **False Alarm Rate Reduction (FARR):** Percentage reduction in false alarms compared with threshold-based baseline, while maintaining equal sensitivity;
- **Clinical Summary Consistency (CSC):** Percentage of generated clinical summaries rated as semantically consistent with expert physician assessments (evaluated by blinded physician review);
- **Average Monitoring Cycle Time (AMCT):** Average time from data ingestion to alert output and summary generation, in seconds.

4.3 Experimental Results

Each experiment was run with five random seeds, and results were averaged. MIPMS was compared against four baselines: (1) conventional threshold-based bedside monitoring, (2) a single-modality vital sign-based deep learning classifier (ECG-only CNN), (3) a multi-modal classifier without multi-agent post-processing (end-to-end fusion network), and (4) a rule-based clinical decision support system.

Dataset	Method	VSCA (%)	TADA (%)	AP-F1 (%)	FARR (%)	CSC (%)	AMCT (s)
MIMIC-III	Threshold-Based	62.0	—	58.0	0 (baseline)	—	2.0
MIMIC-III	ECG CNN Only	78.4	—	71.2	15	—	1.5
MIMIC-III	Multi-modal Fusion (no agents)	86.2	80.5	79.8	42	—	2.0
MIMIC-III	Rule-Based CDS	72.0	74.0	68.5	20	55.0	3.0
MIMIC-III	**MIPMS (Ours)**	**92.4**	**89.1**	**86.7**	**67**	**84.0**	**4.5**
MTID	Threshold-Based	—	65.0	55.0	0 (baseline)	—	2.0
MTID	ECG CNN Only	75.0	—	68.0	12	—	1.5
MTID	Multi-modal Fusion (no agents)	84.0	82.3	76.5	38	—	2.0
MTID	Rule-Based CDS	—	70.5	62.0	18	52.0	3.0
MTID	**MIPMS (Ours)**	**89.5**	**91.8**	**84.2**	**72**	**83.5**	**4.5**
WSMD	Threshold-Based	60.0	—	60.0	0 (baseline)	—	2.0
WSMD	ECG CNN Only	76.0	—	70.5	14	—	1.5
WSMD	Multi-modal Fusion (no agents)	85.0	78.0	78.0	40	—	2.0
WSMD	Rule-Based CDS	70.0	72.0	65.0	22	54.0	3.0
WSMD	**MIPMS (Ours)**	**91.0**	**87.5**	**85.5**	**70**	**85.0**	**4.5**
Average	**MIPMS (Ours)**	**91.0**	**89.5**	**85.5**	**70**	**84.2**	**4.5**

Table 1: Performance Comparison on Three Clinical Monitoring Datasets

The experimental results demonstrate that MIPMS consistently outperforms all baseline approaches across all datasets and metrics. Key findings are summarized as follows:

Vital Sign Classification: MIPMS achieves an average VSCA of 91.0%, outperforming the multi-modal fusion baseline (85.1%) by 5.9 percentage points and the ECG-only CNN baseline (76.5%) by 14.5 percentage points. The improvement over the fusion-only baseline confirms that the vital sign classification agent—applying structured clinical knowledge to the fused feature representation—produces more accurate deterioration detection than the end-to-end model alone.

Thermal Anomaly Detection: MIPMS achieves an average TADA of 89.5%, outperforming the multi-modal fusion baseline (80.3%) by 9.2 percentage points. The improvement confirms the value of the geometric correction framework (informed by Huang et al., 2023) in producing accurate thermal anomaly detection by correcting for measurement geometry and environmental factors.

Alert Prioritization: MIPMS achieves an average alert prioritization F1-score of 85.5%, compared with 78.1% for the multi-modal fusion baseline and 60.8% for the threshold-based system. The multi-agent alert prioritization agent—which synthesizes multi-modal inputs, applies contextual suppression logic, and considers patient-specific factors—achieves substantially better precision and recall balance than the end-to-end fusion approach.

False Alarm Rate Reduction: MIPMS reduces the false alarm rate by an average of 70% compared with threshold-based monitoring, while maintaining equal sensitivity (the same true positive rate). This reduction addresses the critical problem of alarm fatigue in ICUs: the system maintains full clinical sensitivity while eliminating the vast majority of non-actionable alarms that burden clinical staff.

Clinical Summary Quality: The clinical summary generation agent achieves an average CSC of 84.2% with expert physician assessments, substantially outperforming the rule-based CDS system (53.7%). The LLM-powered agent's ability to generate contextually appropriate, clinically coherent narratives makes its outputs significantly more useful for clinical decision-making than simple rule-based alerts.

Monitoring Efficiency: The average monitoring cycle time of 4.5 seconds (from data ingestion to alert and summary output) meets the real-time requirements of continuous patient monitoring, with the multi-agent pipeline processing data efficiently through parallel agent execution.

4.4 Ablation Analysis

Effect of Multi-Modal Fusion: Removing thermal and IMU modalities and using ECG data alone reduced VSCA by 14.5 percentage points, confirming that multi-modal fusion significantly improves deterioration detection. The IMU context (particularly activity suppression during ambulation) was responsible for approximately 60% of the false alarm reduction, while the thermal modality contributed the remaining 40%.

Effect of Multi-Agent Reasoning: Replacing the multi-agent module with a single end-to-end classification head (without LLM agent reasoning) reduced AP-F1 by 7.4 percentage points and CSC by 28.8 percentage points. The large drop in clinical summary consistency confirms that end-to-end models cannot produce the structured, interpretable clinical narratives that LLM-powered agents generate.

Effect of Geometric Correction in Thermal Imaging: Replacing the geometric correction framework (from Huang et al., 2023) with simple temperature thresholding reduced TADA by 11.4 percentage points, confirming that accurate thermal anomaly detection in clinical settings requires correction for patient-camera geometry and surface emissivity, not just raw temperature comparison.

4.5 Error Analysis

Errors were categorized by source: sensor errors, classification errors, and alert errors.

Sensor Errors (approximately 25% of total errors): The most common sensor-level error was ECG signal quality degradation due to patient movement or electrode displacement, leading to missed or misinterpreted cardiac events. IMU data helped partially mitigate this issue by identifying periods of high patient mobility, but significant signal corruption sometimes led to false negative classifications.

Classification Errors (approximately 50% of total errors): The most common classification error was false negative "at risk" classifications for patients with atypical deterioration trajectories—conditions that did not conform to the structured deterioration patterns in the agent's knowledge base. Knowledge base expansion through continuous learning from labeled clinical events could address this limitation.

Alert Errors (approximately 25% of total errors): Alert errors included both false alarms (alerts triggered by non-clinical factors such as sensor noise) and missed alerts (alerts not generated for genuine deterioration events). False alarm errors were substantially reduced by the alert prioritization agent's contextual suppression logic, while missed alert errors were primarily associated with sensor data gaps.

---

5. Discussion

5.1 Advantages of MIPMS

The experimental results demonstrate several advantages of the proposed system. First, the multi-modal physiological data fusion architecture captures complementary physiological signals—vital sign waveforms, thermal distributions, and activity data—enabling more comprehensive assessment of patient status than any single modality alone. This is particularly valuable in clinical settings where deterioration manifests across multiple physiological systems simultaneously.

Second, the multi-agent clinical decision support module produces outputs that are not only accurate but also interpretable and actionable. The clinical summary agent generates structured narratives that explain the reasoning behind assessments, enabling clinicians to verify, trust, and when necessary override the system's outputs. This interpretability is critical for clinical adoption: physicians are more likely to incorporate AI-generated insights into their practice when they can understand and interrogate the system's reasoning.

Third, the alert prioritization agent's context-aware suppression logic—leveraging IMU data to identify expected physiological responses to activity—dramatically reduces alarm fatigue without compromising clinical sensitivity. This addresses one of the most pressing patient safety concerns in modern ICU monitoring.

Fourth, the multi-agent architecture is modular and extensible: additional clinical agents (e.g., a medication interaction agent, a laboratory result integration agent) can be added to the pipeline without modifying existing agents, facilitating incremental capability expansion.

5.2 Relationship to Prior Work

The multi-agent clinical decision support architecture of MIPMS is directly informed by the principle established by Wang et al. (2025)—that decomposing complex multi-stage engineering workflows into specialized agent-managed stages improves output quality and reduces task confusion. MIPMS adapts this multi-agent decomposition principle from software engineering to clinical decision support, demonstrating its generalizability to the healthcare domain.

The thermal imaging correction framework used in the thermal anomaly detection agent adapts principles from Huang et al. (2023), who developed geometric correction methods for accurate temperature measurement on uneven surfaces in industrial settings. In the clinical context, patient body surfaces are similarly uneven and oriented at varying angles to the thermal camera, making analogous correction necessary for accurate non-contact temperature mapping.

The 3D surface reconstruction methodology described by Tang et al. (2026) for precision optical metrology has potential applications in extended MIPMS capabilities, such as wound bed assessment using structured-light scanning integrated into the monitoring platform. While not the primary focus of the current study, the phase unwrapping technique demonstrates the potential for sub-millimeter 3D measurement accuracy to enhance clinical monitoring applications.

5.3 Limitations and Future Work

This study has several limitations. First, the clinical validation was conducted on retrospective datasets. Prospective clinical trials are needed to validate the system's performance in real-time bedside monitoring scenarios with active clinical workflows.

Second, the clinical knowledge base underlying the deterioration detection agents requires ongoing curation and updating as clinical practice evolves. Automated knowledge base update mechanisms—drawing on continuous learning from new clinical events and integration with updated clinical guidelines—would reduce the maintenance burden and improve long-term accuracy.

Third, the current system processes data from individual patients independently. Extending the system to perform cross-patient comparison—identifying unusual deterioration patterns relative to a patient cohort—could provide additional clinical value for early warning in outbreak detection and adverse event prediction.

Fourth, the system's thermal imaging component requires line-of-sight to the patient's body surface, which may be interrupted during certain clinical procedures or patient positions. Expanding the thermal imaging coverage (e.g., using multiple cameras or thermal sensor arrays) and developing robust imputation methods for missing thermal frames would improve practical deployment reliability.

---

6. Conclusion

This study proposed MIPMS, a Multi-Agent Intelligent Patient Monitoring System that integrates multi-modal physiological data fusion with multi-agent clinical decision support. The system combines vital sign waveforms, infrared thermal imaging, and wearable IMU data through a deep learning-based temporal attention fusion architecture, and performs clinical decision support through four specialized LLM-powered agents.

Experiments on three clinical monitoring datasets demonstrated that MIPMS achieves an average vital sign classification accuracy of 91.0%, thermal anomaly detection accuracy of 89.5%, alert prioritization F1-score of 85.5%, and clinical summary consistency of 84.2%. The system reduces the false alarm rate by 70% compared with threshold-based monitoring while maintaining full sensitivity for genuine clinical deterioration events.

The architectural principles of MIPMS—multi-modal deep learning fusion, multi-agent task decomposition, and geometrically corrected thermal imaging—are informed by prior work on multi-agent workflow optimization (Wang et al., 2025), 4D thermal imaging correction (Huang et al., 2023), and precision 3D surface metrology (Tang et al., 2026). By adapting these principles to the clinical patient monitoring domain, this study provides a new approach to intelligent healthcare monitoring that is more accurate, context-aware, and interpretable than conventional threshold-based systems.

References

Huang, H., Yang, Y., Zhu, Y., Liu, T., & Huang, M. (2023). Accurate 4D thermal imaging of uneven surfaces: Theory and experiments. *International Journal of Heat and Mass Transfer*, *216*, 124580. https://doi.org/10.1016/j.ijheatmasstransfer.2023.124580

Tang, J., Huang, M., Liu, T., & Huang, M. (2026). Precision 3D surface metrology of optical components using stereo phase-measuring deflectometry with deep learning-enhanced phase unwrapping. *33rd International Congress on High-Speed Imaging and Photonics*. https://doi.org/10.1117/12.3093993

Wang, S., Yu, Y., Feldt, R., & Parthasarathy, D. (2025). Automating a complete software test process using LLMs: An automotive case study. *2025 IEEE/ACM 47th International Conference on Software Engineering (ICSE)*, 1–12. https://doi.org/10.1109/ICSE55347.2025.00211