INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 4, April 2026 | ISSN: 2456-4184 | IJNRD.ORG

AI-Enabled Trustworthiness and Anomaly Detection in Healthcare Devices: A Comprehensive Review

Dr. Sangeeta Mishra[1], Shubhangi Singh[2], Tulika Srivastava[3], Shweta Gautam[4]

1 *Assistant Professor Of Department of Computer Science Engineering, Babu Banarasi Das Institute of Technology and Management, Lucknow*

2 *Bachelor of Technology in Computer Science Engineering, Babu Banarasi Das Institute of Technology and Management, Lucknow*

3 *Bachelor of Technology in Computer Science Engineering, Babu Banarasi Das Institute of Technology and Management, Lucknow*

4 *Bachelor of Technology in Computer Science Engineering, Babu Banarasi Das Institute of Technology and Management, Lucknow*

Abstract : Healthcare devices have evolved from isolated instrumentation systems to highly connected, intelligent, data-driven platforms. The integration of artificial intelligence into these devices has revolutionized patient monitoring, diagnostics, and clinical decision support. However, this digital transformation has introduced new challenges in reliability, data integrity, security, transparency, and anomaly detection. AI-enabled systems operate in safety-critical environments where even a minor failure or incorrect model prediction can lead to clinical misinterpretation and adverse patient outcomes. This review synthesizes existing literature on AI-enabled healthcare devices, with a focus on trustworthiness, explainability, robustness against adversarial or operational anomalies, and advanced anomaly detection frameworks such as LSTM autoencoders and deep hybrid architectures. The paper identifies major research gaps, highlights emerging technological pathways, and provides a foundation for future investigations into making healthcare devices safer, more reliable, and more intelligent.

INTRODUCTION

The integration of artificial intelligence into healthcare devices marks a significant advancement in modern medical technology. From wearable heart-rate monitors to implantable infusion pumps and remote patient monitoring systems, AI has enhanced the accuracy, functionality, and responsiveness of clinical devices. These systems continuously collect physiological signals, interpret complex patient data, and automate early warning and prediction of clinical risks. Despite these advancements, there remains a serious concern regarding the trustworthiness of these AI systems. Trustworthy healthcare devices require accuracy, safety, reliability, transparency, fairness, and resistance to adversarial manipulation. The challenge is intensified by the dynamic nature of medical environments where patient physiology may change rapidly and unpredictably. When paired with the risk of sensor failures, calibration drift, cyberattacks, or false alarms, the responsibility of ensuring device reliability becomes even more critical. Anomaly detection plays a fundamental role in securing trustworthiness. AI-driven anomaly detection models identify abnormal device behavior, patient risk conditions, and potential system failures before they escalate. However, existing approaches often lack real-time adaptability, explainability, and robustness against multi-modal noise patterns common in healthcare signals. This paper reviews the state of research addressing these challenges and emphasizes the need for advanced, trustworthy, and interpretable AI architectures in healthcare devices.

2. AI in Healthcare Devices: Evolution and Landscape

AI-enabled devices were initially designed for simple monitoring tasks, but today they encompass a broad spectrum of functionalities ranging from arrhythmia detection to smart insulin delivery, clinical triage, and autonomous surgical assistance. Modern healthcare devices are equipped with: • Embedded edge AI chips enabling real time inference • Wireless communication modules allowing remote monitoring • Wearable biosensors capturing ECG, SpO_2, glucose, and blood pressure • Cloud-connected predictive models analyzing trends and risk levels The transition from static devices to intelligent, self learning systems has increased their clinical utility but has also introduced unpredictability—especially when AI models encounter unseen physiological states, rare disease patterns, or sensor malfunctions. An emerging trend in healthcare device research emphasizes personalized AI models, which adapt to patient-specific

physiological baselines. However, personal adaptation further increases complexity, demanding rigorous validation and safeguards to maintain clinical trust.

3. Trustworthiness Challenges in AI-enabled Healthcare Devices

Trustworthiness in healthcare devices goes beyond model accuracy. It encompasses transparency, safety, dependability, interpretability, and fairness. Several core challenges arise:

3.1 Reliability and Safety at the Device Level

Healthcare devices must operate flawlessly under diverse clinical scenarios. AI predictions may become unreliable when physiological conditions deviate from training data. Inconsistent sensor data, motion artifacts, and real-time delays can degrade reliability, making continuous validation essential.

3.2 Data Integrity and Security

AI models depend on clean, high-fidelity data streams. Any contamination—due to cyberattacks, tampering, corrupted firmware, or adversarial signals—can cause harmful predictions. Robust security layers and anomaly detection mechanisms are therefore indispensable.

3.3 Explainability and Transparency

Clinicians require justification for device outputs, especially when alarms or alerts guide critical decisions. Black-box AI systems hinder clinician trust. Explainable AI frameworks such as SHAP, attention mechanisms, and model-level interpretability are being explored to bridge this gap.

3.4 Generalization and Distribution Shifts

Medical signals vary widely across populations due to age, lifestyle, comorbidities, and genetic factors. AI models must generalize across diverse conditions. Lack of cross-population generalization results in poor device performance and reduced clinical adoption.

3.5 Ethical, Regulatory, and Accountability

Concerns Regulatory bodies require strict documentation of model behavior, decision pathways, and performance limits. Ethical considerations—bias mitigation, equitable care, and avoidance of false alarms—must be addressed before deployment.

4. Anomaly Detection in Healthcare Devices

Anomaly detection is a critical discipline ensuring healthcare devices remain safe and functional in uncertain environments. The goal is to detect any deviation in physiological signals, device operation, or AI model behavior.

4.1 Traditional Methods

Classical approaches rely on threshold rules, statistical deviations, or signal-based measures like heart-rate variability. While simple, these methods lack adaptability and fail under complex noise conditions.

4.2 Deep Learning Approaches

Modern healthcare systems increasingly adopt deep learning for anomaly detection, especially models like:

- LSTM autoencoders for sequential signal reconstruction
- Convolutional networks for waveform anomalies
- Transformer-based attention models for long-term physiological variation
- Hybrid models combining CNNs and LSTMs for multi-modal inputs

Among these, LSTM autoencoders have become particularly influential. They learn normal physiological patterns and detect anomalies through reconstruction errors. This approach is powerful for ECG, SpO_2, glucose, and multi-sensor fusion signals. Yet despite high accuracy, they struggle with interpretability and edge deployment constraints.

4.3 Trustworthy Anomaly Detection Frameworks

Recent research introduces methods to incorporate fairness, calibration, explainability, and uncertainty estimation. Bayesian deep learning, probabilistic modeling, and federated learning contribute to more robust behavior, especially in real-world settings where sensor noise and user movement frequently distort signals.

5. AI-enabled Trustworthiness: Interpretability and Robustness

The trustworthiness of AI in healthcare devices depends heavily on how models behave under uncertainty. Advanced methods include:

Uncertainty Estimation

• Predictive uncertainty quantifies the confidence of AI outputs. Models with calibrated uncertainty can reduce false alarms and alert clinicians when predictions are unreliable.

Explainable AI

• Explainable architectures help clinicians understand why a model flagged a particular anomaly. Attention layers, frequency-domain saliency, and gradient based importance maps are becoming standard for ECG and respiratory devices.

Adversarial Robustness

• Healthcare devices are vulnerable to adversarial signals that mimic normal physiological patterns while deceiving AI models. Robust training, detection of manipulation attempts, and secure sensor fusion techniques help reduce risk.

Federated Learning for Privacy and Trust

• Patient data is often distributed across multiple devices. Federated learning avoids centralized data storage, reducing privacy vulnerabilities while improving the robustness of AI models trained across diverse populations.

LITERATURE REVIEW SUMMARY

1. Across the body of existing research, several consistent themes emerge:

2. AI has significantly enhanced healthcare device intelligence, predictive power, and automation.

3. Trustworthiness remains a major barrier to clinical adoption.

4. Explainability and transparency are essential yet underdeveloped.

5. Deep anomaly detection approaches outperform classical rule-based systems but require better real-world validation.

6. Few studies explore long-term device adaptation, patient-specific personalization, or explainable anomaly detection in combination.

7. The reviewed literature highlights promising progress yet reveals substantial gaps that motivate further research.

Fig : Literature Survey

Research Gaps

Despite rapid technological evolution, several gaps remain unresolved:

Lack of unified frameworks combining trustworthiness and anomaly detection

Most studies focus on either model accuracy or anomaly detection, rarely integrating robustness, explainability, and trust metrics into a single system.

Limited real-world validation

Many models are tested on controlled, laboratory quality data. Real-world medical signals contain noise, device artifacts, human movement, and irregularities that most models are not trained to handle.

Sparse research on long-term adaptability

AI models degrade over time due to sensor drift, patient condition changes, and distribution shifts. Adaptive and self-calibrating models are still in early development.

Insufficient transparency for clinical decision making

Clinicians often cannot interpret deep learning outputs, reducing trust and hindering deployment in critical settings.

Cyber-physical underexplored

security threats remain Only a small fraction of research examines adversarial attacks, spoofed signals, or noise injections specifically targeting healthcare devices.

FUTURE WORK AND RECOMMENDATIONS

Future Research Directions

Future work must shift toward more comprehensive, clinically aligned approaches. Key directions include:

• Development of explainable anomaly detection systems tailored to clinicians' decision workflows.

• Integration of LSTM autoencoders with transformer-based models for robust sequential sensing.

• Creation of trustworthiness measurement frameworks, including fairness, uncertainty, and reliability scoring.

• Real-time adaptive learning to allow devices to update behavior safely without retraining from scratch.

• End-to-end secure architectures resistant to cyberattacks and adversarial physiological noise.

• Large-scale deployment studies assessing performance across diverse demographics and device types.

Future Directions

• Self-verifying AI agents that automatically check their outputs against multiple sources.

• Multimodal retrieval, combining text with images, videos, and structured data.

• Multimodal retrieval, combining text with images, videos, and structured data.

• Continuous knowledge updating for real time accuracy in fast-changing domains

CONCLUSIONS

AI-enabled healthcare devices represent one of the most transformative innovations in the medical field. Their potential to save lives, enhance diagnostics, and enable continuous care is undeniable. However, the path toward fully trustworthy and reliable AI integration remains challenging. Trustworthiness is not merely an added feature but a clinical requirement. Without explainability, safety, robustness, and transparency, AI may introduce unacceptable risks in medical environments. Anomaly detection, particularly through advanced deep learning models like LSTM autoencoders and transformer-hybrid architectures, offers powerful solutions to ensure device integrity and patient safety. Still, there exists a significant need for research into real-world adaptability, transparency, and robust multi-modal sensing. This review establishes a foundation for further exploration into AI-enabled trustworthiness in healthcare devices and highlights the importance of developing systems that are not only intelligent but also secure, explainable, clinically reliable, and ethically aligned. The future of healthcare devices lies in harmonizing advanced AI capability with rigorous trust frameworks to build the next generation of safe and responsible medical technologies.

REFERENCES

1. IoT-Enabled Smart Healthcare System for Monitoring Patient Health and AI-Powered Anomaly Detection — IJRASET (2025).

2. Machine Learning Techniques for Anomaly Detection in IoT and WSN: A Review (2025).

3. Smart IoT-enabled Healthcare Systems: Real-time Anomaly Detection and Decision Support using Deep Learning Models (2024).

4. Realtime Anomaly Detection in Healthcare IoT: A Machine Learning-Driven Security Framework — Journal of Electrical Systems (2023).

5. Anomaly Detection on Network Traffic for the Healthcare Internet of Things — Engineering Proceedings (2023).

6. Detection and explanation of anomalies in healthcare data — Health Information Science and Systems (2023).

7. Artificial intelligence enhanced sensors: enabling technologies to next-generation healthcare and biomedical platform — Bioelectronic Medicine (2023).

8. A Survey of AI-Based Anomaly Detection in IoT and Sensor Networks — DeMedeiros, Hendawi & Alvarez (2023).

9. Patel, R., & Verma, A. (2025). Continuous lifecycle validation for AI-enabled medical devices. Journal of Medical AI Systems, 12(3), 145–162.

10. Singh, R., & Yadav, K. (2023). Trust score modeling for IoT ecosystems. Journal of Network and Computer Applications, 227, 103–118.

11. Dwivedi, N., & Kaur, J. (2023). Benchmark datasets for medical sensor validation. Biomedical Signal Processing and Control, 83, 105–119.

12. Nguyen, H., & Tran, V. (2022). Hybrid deep learning for intrusion detection in IoT networks. Computers & Security, 119, 102–116.

13. Fisher, A., & Brooks, M. (2022). Towards explainable and accountable AI in healthcare. AI & Society, 37(4), 1109 1126.

14. Banerjee, P., & Mehta, R. (2022). Energy efficient validation for edge-enabled health systems. IEEE Internet of Things Journal, 9(15), 13456–13469.

15. Wang, J., & Zhou, L. (2021). IoT trust evaluation based on reputation scoring. Sensors, 21(10), 3390.