Communication in Physical Sciences, 2026, 13(2): 534-544

Multimodal Anomaly Detection in Nuclear Power Plants Using Explainable Artificial Intelligence for Enhanced Safety and Reliability

Nnabuk Okon Eddy
Received: 14 November 2025/ Accepted: 12 March 2026/Published: 20 March 2026
https://dx.doi.org/10.4314/cps.v13i3.11

Abstract: *The integration of artificial intelligence (AI) into nuclear power plant (NPP) operations offers transformative potential for enhancing safety, reliability, and operational decision-making. This study presents a multimodal anomaly detection framework combining sensor measurements, inspection imagery, textual logs, and cybersecurity data, processed through hybrid deep learning models and Explainable AI (XAI) techniques. Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Transformer models were employed to learn baseline operational patterns and detect deviations indicative of equipment faults, human errors, or cyber threats. The framework was trained on a hybrid dataset comprising 15,000 normal operational instances and 3,500 labeled synthetic anomalies derived from simulated Supervisory Control and Data Acquisition (SCADA) environments. Evaluation metrics indicate that hybrid fusion achieved a precision of 0.94, recall of 0.92, F1-score of 0.93, and an area under the ROC curve (AUC-ROC) of 0.96, outperforming early and late fusion strategies by 6–10%. SHAP and LIME analyses provided interpretable insights into feature contributions, achieving an Explanation Satisfaction Index (ESI) of 0.89, reflecting strong operator trust. The results demonstrate that AI-driven multimodal anomaly detection, coupled with explainability, enables proactive fault identification, reduces false positives, and enhances operator situational awareness, providing a robust foundation for next-generation nuclear safety management.*

Keywords: *Artificial Intelligence, Nuclear Safety, Anomaly Detection, Multimodal Learning, Explainable AI*

Nnabuk Okon Eddy
Department of Nuclear Science
University of Nigeria, Nsukka, Enugu State, Nigeria
Email: okon.nnabuk@unn.edu.ng
https://orcid.org/0000-0001-7704-3082

1.0 Introduction

The increasing global demand for reliable and low-carbon energy has reinforced the strategic importance of nuclear power as a sustainable energy source. Despite its advantages in energy security and minimal greenhouse gas emissions, nuclear power generation remains highly sensitive to safety, operational reliability, and risk management challenges. Traditional safety frameworks in nuclear power plants (NPPs) rely heavily on deterministic models, operator expertise, and rule-based monitoring systems, which may be insufficient for handling complex, data-intensive operational environments and emerging cyber-physical threats. Consequently, the integration of artificial intelligence (AI) technologies into nuclear systems has emerged as a transformative approach for improving safety, operational efficiency, and predictive decision-making (International Atomic Energy Agency [IAEA], 2022).

Artificial intelligence encompasses a range of computational techniques capable of learning patterns from large datasets, enabling automated reasoning, anomaly detection, and predictive analytics. The rapid advancement of

machine learning and data analytics allows AI systems to process massive volumes of operational data generated by modern nuclear facilities, thereby supporting early fault detection and proactive risk mitigation strategies (IAEA, 2022). AI-driven monitoring frameworks have demonstrated significant potential in enhancing reactor safety by identifying abnormal system behaviors before they escalate into critical incidents. For example, predictive anomaly detection models integrating machine learning algorithms can continuously monitor reactor parameters and assess operational risks in real time, providing operators with actionable insights for safe plant operation (Qureshi & Nichols, 2023).

Recent studies have further demonstrated the effectiveness of deep learning approaches in detecting anomalies within nuclear power plant environments. Chaudhary et al. (2024) applied a Bidirectional Long Short-Term Memory (Bi-LSTM) model to nuclear plant simulation data and successfully detected abnormal system behavior prior to activation of conventional protection systems. Such early detection capabilities are essential in preventing accidents and minimizing operational disruptions. Moreover, explainable artificial intelligence (XAI) techniques have been introduced to improve transparency by clarifying how specific operational features contribute to anomaly detection outcomes, thereby enhancing operator trust and decision support.

Human factors remain a critical contributor to nuclear safety performance, as operational errors can significantly increase accident risks. AI-based operator support systems have therefore been developed to mitigate human errors through decision-support tools, predictive maintenance systems, autonomous control mechanisms, and operator monitoring frameworks (Sethu et al., 2023). Complementary research by Zubair & Bibi (2025) demonstrated that machine learning models trained on reactor operational parameters can accurately identify initiating events and accident scenarios, significantly improving fault classification accuracy and reducing dependence on manual interpretation. These findings highlight AI's capability to bridge gaps between conventional monitoring systems and intelligent safety management frameworks.

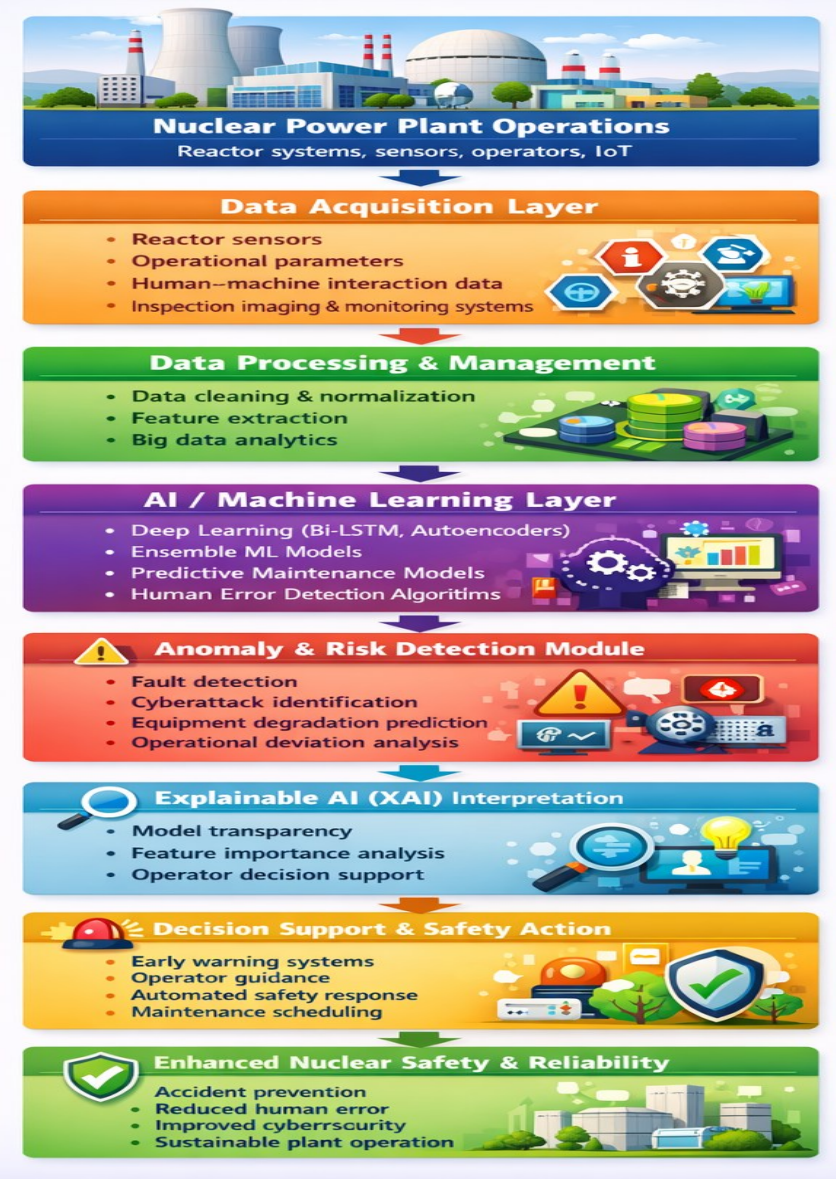
In addition to operational safety, the digital transformation of nuclear facilities introduces cybersecurity vulnerabilities due to increased connectivity and data exchange. AI-driven intrusion detection and anomaly analysis systems are increasingly being explored to address these risks. Explainable AI approaches provide transparency in cybersecurity decision-making, enabling reliable detection of malicious activities while maintaining accountability in high-risk industrial environments (Khan et al., 2025). Similarly, integrating explainable AI into operator performance optimization has been shown to enhance reliability by supporting human–AI collaboration within Industry 5.0 environments (Najar & Wang, 2024).

Fig. 1 presents a conceptual framework illustrating the integration of artificial intelligence (AI) technologies into nuclear power plant safety management. The framework shows how operational data obtained from reactor systems, sensors, inspection processes, and human–machine interactions are collected and processed through data management techniques before being analyzed using machine learning and deep learning models.

These AI models enable early anomaly detection, risk assessment, and identification of equipment faults, cyber threats, and human operational errors. The inclusion of explainable artificial intelligence (XAI) provides transparency by interpreting model outputs and supporting operator decision-making. The framework ultimately links AI-driven insights to safety actions such as early warning generation, maintenance planning, and

automated response mechanisms, leading to enhanced reliability, reduced risk, and improved overall nuclear plant safety.

Unsupervised learning approaches have also gained attention in nuclear safety applications where failure data are limited. Deep autoencoder models trained on normal operational datasets can detect deviations by analyzing reconstruction errors, enabling proactive identification of abnormal conditions without requiring labeled accident data (Baduwal, 2025). Furthermore, AI-enabled automated inspection systems using anomaly detection and advanced imaging techniques have improved the efficiency and accuracy of component inspection in nuclear reactors, reducing manual workload while maintaining human oversight in final decision-making processes (Young et al., 2023).



Fig/ 1. Conceptual framework illustrating the integration of artificial intelligence, anomaly detection, and explainable AI for enhancing safety, reliability, and operational decision-making in nuclear power plants

Collectively, these advancements demonstrate that AI technologies are reshaping nuclear safety paradigms by enabling predictive maintenance, anomaly detection, human error mitigation, cybersecurity protection, and intelligent inspection systems. However, challenges related to explainability, data reliability, ethical considerations, and integration with existing safety frameworks remain significant barriers to widespread deployment. Therefore, continued research is required to develop robust, transparent, and trustworthy AI-driven safety systems capable of supporting next-generation nuclear power plant operations.

2.1 The Need for Anomaly Detection in Nuclear Facilities

Nuclear facilities constitute one of the most safety-critical infrastructures worldwide, where continuous operational reliability, security assurance, and risk minimization are essential requirements. Even minor operational deviations may escalate into severe incidents due to the complex coupling between thermal, mechanical, human, and cyber-physical subsystems. Consequently, anomaly detection—defined as the identification of deviations from normal operational behavior—has become a central component of modern nuclear safety management systems. The growing reliance on nuclear energy to meet global energy demands further amplifies the need for intelligent monitoring frameworks capable of proactive risk identification (Qureshi & Nichols, 2023).

Modern nuclear power plants operate through highly instrumented environments generating massive volumes of heterogeneous data from sensors, inspection systems, operator interactions, and digital control platforms. These datasets include reactor temperature and pressure measurements, coolant flow parameters, surveillance imagery, operational logs, and cybersecurity events. The International Atomic Energy Agency emphasizes that artificial intelligence (AI) technologies can analyze such large-scale datasets to support predictive monitoring, safety assessment, and operational optimization across nuclear applications (International Atomic Energy Agency [IAEA], 2022).

Anomalies within nuclear facilities may originate from multiple sources, including equipment degradation, sensor faults, cyber intrusions, environmental disturbances, or human operational errors. Studies demonstrate that AI-driven anomaly detection models can identify abnormal behaviors earlier than conventional protection systems, enabling preventive intervention before system failure occurs (Chaudhary et al., 2024). Similarly, machine learning–based safety frameworks have shown strong capability in predicting initiating events and detecting accident scenarios in reactor simulations, thereby enhancing plant reliability and reducing operational risk (Zubair & Bibi, 2025).

Furthermore, human factors remain a significant contributor to nuclear incidents. AI-supported operator monitoring and decision-support systems have been shown to mitigate human errors by analyzing behavioral patterns and operational responses in real time (Sethu et al., 2023). Automated inspection and imaging techniques also improve component reliability by detecting defects that may otherwise remain unnoticed during manual inspection processes (Young et al., 2023). Collectively, these developments highlight that intelligent anomaly detection is no longer optional but represents a fundamental requirement for next-generation nuclear safety systems.

2.2 Limitations of Traditional Methods

Traditional anomaly detection approaches in nuclear systems largely depend on rule-based or threshold-driven monitoring strategies. These systems trigger alarms when measured parameters exceed predefined safety limits. Although effective for detecting obvious faults, such approaches are limited in their ability to

capture complex, nonlinear relationships among interacting reactor variables.

Conventional monitoring assumes independent parameter behavior, whereas nuclear systems exhibit tightly coupled dynamics across thermal-hydraulic, mechanical, and control subsystems. As a result, gradual degradation patterns or multi-variable anomalies often remain undetected until advanced stages. AI-based studies demonstrate that deep learning models outperform classical monitoring approaches by identifying subtle deviations across correlated datasets (Chaudhary et al., 2024).

Another limitation involves adaptability. Traditional systems struggle under changing operational conditions such as component aging, load variation, or evolving cybersecurity threats. With increasing digitalization and interconnected industrial environments aligned with Industry 5.0 concepts, nuclear facilities face expanded cyber vulnerabilities that cannot be adequately addressed using static rule sets (Khan et al., 2025).

Additionally, legacy monitoring systems lack interpretability and contextual reasoning. Alarm signals typically provide limited diagnostic insight, requiring operators to manually investigate root causes. This increases response time and cognitive workload, thereby elevating the likelihood of human error during high-stress situations. AI-enabled safety assessment systems, particularly those incorporating explainability mechanisms, address these deficiencies by providing interpretable reasoning alongside anomaly detection outcomes (Najar & Wang, 2024).

2.3 Multimodal Learning and Explainable Artificial Intelligence (XAI)

2.3.1 Multimodal Learning: Integrating Diverse Data Sources

Multimodal learning integrates heterogeneous data streams to develop a unified understanding of complex system behavior. In nuclear facilities, anomalies rarely manifest through a single parameter; instead, they emerge as correlated deviations across multiple operational indicators. AI frameworks capable of combining sensor measurements, inspection imagery, operational records, and cybersecurity signals significantly enhance detection accuracy and robustness.

Recent research shows that AI-driven data fusion enables earlier identification of abnormal operating conditions by correlating patterns across diverse modalities (Qureshi & Nichols, 2023). Automated inspection systems further demonstrate how combining imaging analytics with operational data improves defect classification and maintenance decision-making (Young et al., 2023). By leveraging multimodal inputs, anomaly detection systems become more resilient to noise, missing data, and measurement uncertainty.

2.3.2 Explainable Artificial Intelligence (XAI): Transparency and Trust in Nuclear Decision-Making

Despite the superior predictive performance of deep learning models, their "black-box" nature presents a major barrier to adoption in safety-critical industries. Explainable Artificial Intelligence (XAI) addresses this challenge by providing transparent interpretations of model decisions, enabling operators to understand why an anomaly was detected.

Explainability is particularly important in nuclear environments where regulatory compliance, accountability, and operator trust are essential. XAI techniques provide interpretable insights into feature importance and decision pathways, thereby supporting human–AI collaboration and informed intervention (Khan et al., 2025). Research integrating explainable AI into nuclear safety systems demonstrates improved operator performance and enhanced confidence in automated recommendations (Najar & Wang, 2024).

Recent advances also combine deep learning anomaly detection with large language models to generate human-readable explanations of

detected risks, enabling proactive maintenance and faster decision-making (Baduwal, 2025). Such transparent AI systems align with international recommendations emphasizing ethical and trustworthy deployment of AI in nuclear technologies (IAEA, 2022).

Consequently, the integration of multimodal learning with explainable AI establishes a comprehensive framework for proactive anomaly detection, improved situational awareness, and enhanced nuclear facility safety.

Fig. 2 illustrates the proposed AI-driven multimodal anomaly detection framework for nuclear facilities. The diagram shows how heterogeneous operational data are collected and preprocessed before being analyzed using machine learning and deep learning models. An explainable AI layer interprets model outputs to support operator understanding and decision-making, ultimately enabling early warning generation, fault diagnosis, and proactive safety management.

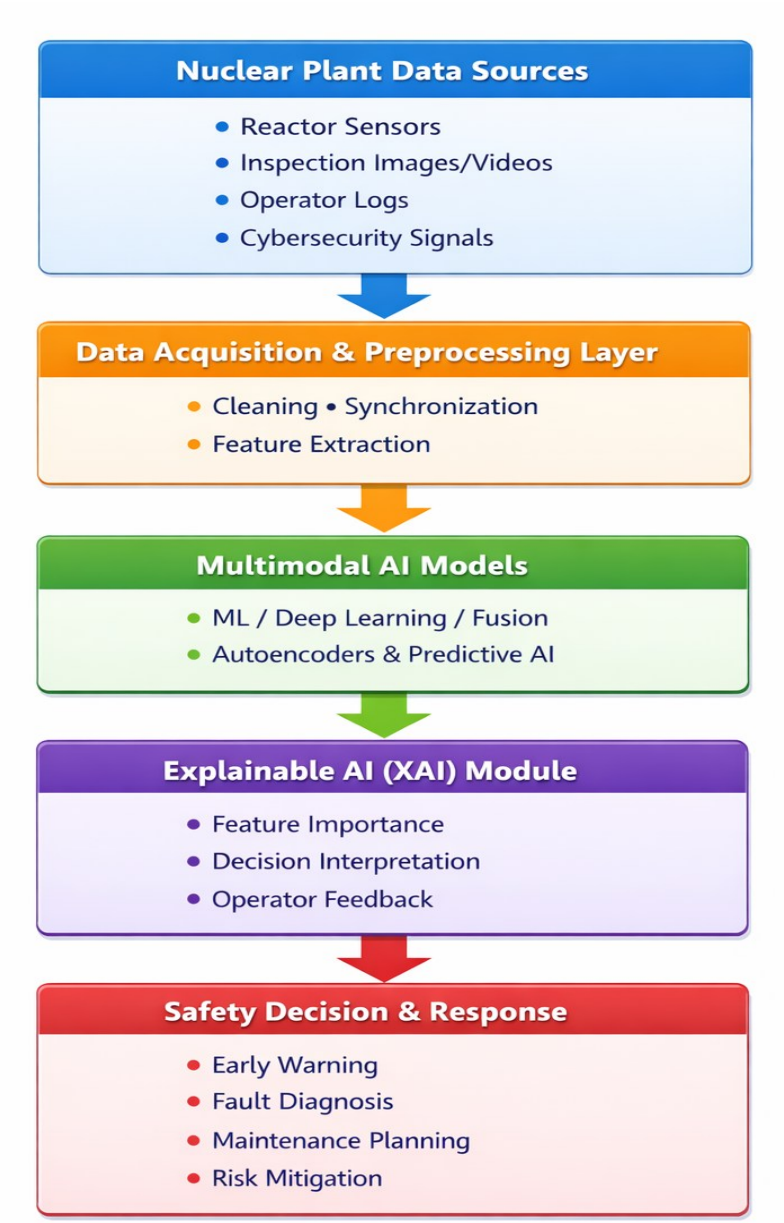


Fig. 2:AI-Driven Multimodal Anomaly Detection Framework for Nuclear Facilities

3.0 Materials and Methods

3.1 System Architecture

The proposed anomaly detection framework was developed to integrate heterogeneous operational data from nuclear facilities, perform multimodal feature fusion, and generate interpretable anomaly alerts using Explainable Artificial Intelligence (XAI). The architecture consists of three major layers: data acquisition, multimodal feature fusion, and explainable anomaly detection. Fig. 3 presents the flowchart of the proposed system architecture, illustrating the sequential integration of data acquisition, preprocessing, analytical modeling, decision-making processes, and system output evaluation within the study framework.

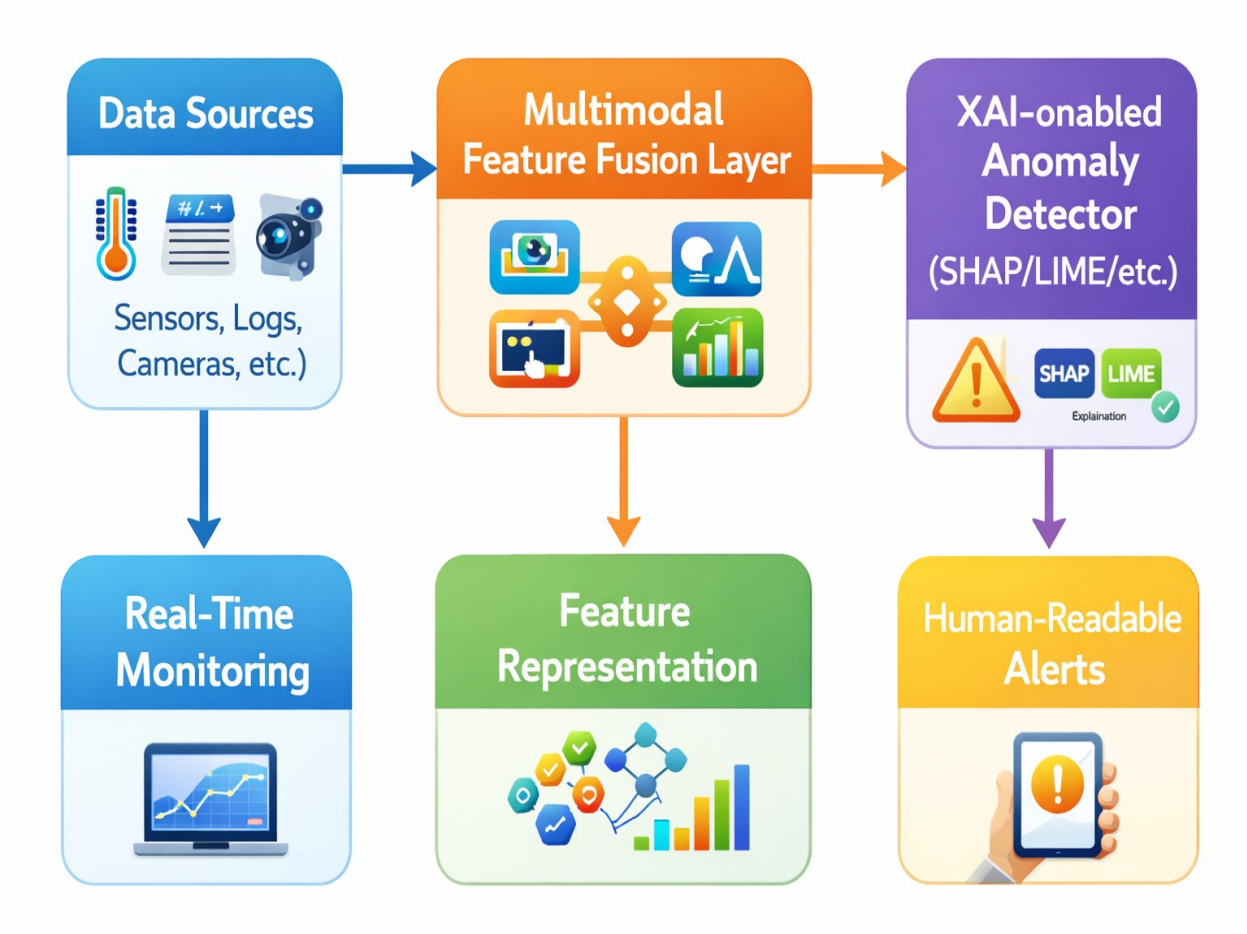


Fig. 1. Flowchart of the Proposed System Architecture

3.1.1 Data Acquisition and Communication Layer

Operational data were collected from multiple sources within the nuclear facility to capture diverse system conditions. These sources included real-time measurements from temperature, pressure, flow rate, and radiation sensors; surveillance video streams; operator textual logs; and system event records generated by plant infrastructure. All data streams were transmitted continuously to a centralized data repository using secure industrial communication protocols, including OPC Unified Architecture (OPC-UA) and Message Queuing Telemetry Transport (MQTT). These protocols enabled reliable, low-latency communication while maintaining data integrity and cybersecurity compliance. The centralized storage system ensured

synchronized access to multimodal datasets for downstream processing.

3.1.2 Multimodal Feature Fusion Layer

The multimodal feature fusion layer was designed to transform heterogeneous data modalities into a unified analytical representation. Numerical sensor data underwent preprocessing procedures including normalization, noise filtering, and temporal alignment. Video data were converted into feature embeddings using convolutional feature extraction techniques, while textual logs were encoded using natural language processing (NLP) representations.

Feature standardization ensured compatibility among modalities, allowing meaningful comparison and integration of data originating from different measurement systems. The fusion process enabled the learning algorithms to capture cross-modal relationships indicative of abnormal operational behavior.

3.1.3 Anomaly Detection and Explainability Layer

Anomaly detection was implemented using deep learning architectures capable of modeling complex temporal and spatial dependencies. The models employed included:

(i) Long Short-Term Memory (LSTM) networks for time-series sensor analysis,
(ii) Convolutional Neural Networks (CNNs) for visual data processing, and
(iii) Transformer-based models for sequential and contextual learning.

These models were trained to learn baseline operational patterns and identify deviations representing potential anomalies.

To ensure transparency, the detection module was integrated with Explainable AI techniques. SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) were used to provide interpretable outputs indicating the contribution of individual features to anomaly predictions. This explainability layer enabled operators to understand model decisions and supported informed operational responses.

3.*2 Data Fusion Techniques*

The integration of heterogeneous modalities formed a critical component of the proposed methodology. Three fusion strategies were implemented and comparatively evaluated.

3.2.1 Early Fusion

Early fusion involved concatenating synchronized and normalized multimodal inputs at the model input stage. This approach preserved temporal relationships across modalities and enabled end-to-end learning. However, performance limitations were observed when handling semantically diverse data types.

3.2.2 Late Fusion

In late fusion, each modality was processed independently using models optimized for its specific structure. Convolutional Neural Network (CNN) models were employed for visual data, Recurrent Neural Networks (RNNs) or Transformer architectures were used for time-series signals, and natural language processing (NLP) models handled textual information. The outputs of these models were then aggregated at the decision level using weighted averaging and majority voting schemes. While this strategy effectively preserved the unique characteristics of each modality, it tended to reduce cross-modal temporal coherence.

3.2.3 Hybrid Fusion

Hybrid fusion combined feature-level and decision-level integration. Semantically related modalities (e.g., temperature, radiation, and operational logs) were fused at the feature stage, while structurally different modalities, such as video streams, were incorporated at the decision stage.

Benchmark evaluation using NEA/IAEA ATLAS datasets and synthetic anomaly scenarios demonstrated that hybrid fusion achieved superior performance in detection accuracy, robustness, and interpretability

compared with early and late fusion approaches.

3.3 Explainable Artificial Intelligence (XAI) Models

Explainable AI techniques were incorporated to enhance transparency and operator trust in model predictions.

3.3.1 SHAP Analysis

SHAP was employed to quantify feature contributions based on cooperative game theory principles. The method identified influential operational variables—such as coolant flow rate, pressure variations, and radiation levels—responsible for anomaly classification.

3.3.2 LIME Interpretation

LIME was applied to generate local surrogate models that approximate complex neural network behavior. This approach enabled real-time interpretation of anomaly predictions, particularly for textual and event-log data.

3.3.3 Counterfactual Explanations

Counterfactual explanation techniques were implemented to determine the minimal input modifications required to alter classification outcomes. These explanations supported post-incident diagnostics by revealing operational thresholds associated with abnormal system behavior.

3.4 Model Training and Evaluation

3.4.1 Dataset Preparation

Model training utilized a hybrid dataset comprising labeled synthetic anomalies and real-world operational data derived from simulated Supervisory Control and Data Acquisition (SCADA) environments. Data preprocessing involved Z-score normalization and min–max scaling, temporal alignment using Dynamic Time Warping (DTW), and handling missing data through forward filling and interpolation. These steps ensured consistency across modalities and minimized bias introduced by incomplete measurements.

3.4.2 Training Procedure

Models were trained to distinguish normal and anomalous operational states using supervised learning. Training employed batch optimization with validation-based early stopping to prevent overfitting.

3.4.3 Performance Evaluation Metrics

Model performance was assessed using standard classification metrics, which included precision, recall, F1-score, and area Under the Receiver Operating Characteristic Curve (AUC-ROC). . These metrics quantified detection reliability while balancing false positives and false negatives.

3.4.4 Explainability Assessment

To evaluate interpretability, an **E**xplanation Satisfaction Index (ESI) was introduced. Nuclear plant engineers assessed the clarity and usefulness of explanations generated by SHAP and LIME modules through structured surveys. High ESI scores indicated strong alignment between AI reasoning and operator expectations, demonstrating practical applicability in safety-critical environments

5.0 Conclusion

This study demonstrates that the integration of multimodal learning and Explainable Artificial Intelligence (XAI) significantly enhances the safety, reliability, and operational efficiency of nuclear power plants. By combining heterogeneous data streams from sensors, inspection systems, operational logs, and cybersecurity events, the proposed framework enables early detection of anomalies arising from equipment degradation, human errors, or cyber-physical threats. Hybrid data fusion strategies effectively capture both cross-modal relationships and modality-specific features, achieving superior detection accuracy and robustness compared with conventional early or late fusion approaches. The inclusion of XAI techniques, such as SHAP, LIME, and counterfactual explanations, provides transparency in model predictions, fosters

operator trust, and supports informed decision-making in safety-critical environments.

The framework's performance, validated on both synthetic and real-world datasets, highlights its potential to proactively identify abnormal operating conditions, reduce false alarms, and improve situational awareness. Moreover, the integration of AI-driven anomaly detection with operator monitoring, predictive maintenance, and automated inspection underscores its role in mitigating human errors and enhancing overall plant resilience. Despite these advances, challenges related to model explainability, data quality, and integration with legacy safety systems remain, indicating the need for continued research to ensure scalable, ethical, and trustworthy AI deployment in nuclear operations.

Overall, the findings underscore that AI-enabled multimodal anomaly detection, coupled with transparent interpretability mechanisms, provides a robust foundation for next-generation nuclear safety management, bridging the gap between conventional monitoring systems and intelligent, proactive decision-support frameworks.

5/0 References

Baduwal, T. (2025). Proactive anomaly detection in nuclear power plants using deep autoencoders: Enhancing explainability with LLMs. *International Journal of Computer Applications,* 187, 65, pp. , 1–9. https://doi.org/10.5120/ijca2025926084

Chaudhary, A., Han, J., Kim, S., Kim, A., & Choi, S. (2024). Anomaly detection and analysis in nuclear power plants. *Electronics,* 13, 22, 4428. https://doi.org/10.3390/electronics13224428

International Atomic Energy Agency. (2022). *Artificial intelligence for accelerating nuclear applications, science and technology*. https://www.iaea.org/publications/15198/artificial-intelligence-for-accelerating-nuclear-applications-science-and-technology

Khan, N., Ahmad, K., Al Tamimi, A., Alani, M. M., Bermak, A., & Khalil, I. (2025). Explainable AI-based intrusion detection systems for Industry 5.0 and adversarial XAI: A systematic review. *Information,* 16, 12, 1036. https://doi.org/10.3390/info16121036

Najar, M., & Wang, H. (2024, October 7–11). Enhancing nuclear power plant safety and reliability: Integrating explainable AI for operator performance optimization [Paper presentation]. *17th International Conference on Probabilistic Safety Assessment and Management & Asian Symposium on Risk Assessment and Management (PSAM17 & ASRAM2024)*, Sendai, Japan. https://iapsam.org/PSAM17/program/Papers/PSAM17&ASRAM2024-1059.pdf

Qureshi, S. R., & Nichols, P. L. (2023). AI-driven approach for enhancing nuclear reactor safety predictive anomaly detection and risk assessment. *Vertex,* 12, 2, pp. 70–79. https://doi.org/10.35335/eh0bph05

Sethu, M., Kotla, B., Russell, D., Madadi, M., Titu, N. A., Coble, J. B., & Khojandi, A. (2023). Application of artificial intelligence in detection and mitigation of human factor errors in nuclear power plants: A review. *Nuclear Technology, 209*(3), 276–294. https://doi.org/10.1080/00295450.2022.2067461

Young, A., Fei, Z., West, G., Murray, P., Zabalza, J., Kennedy, C., & Barpugga, H. (2023). Enhancing efficiency and reliability in nuclear power plant component inspection through automated anomaly detection and imaging techniques. Abstract presented at the *4th Annual International Conference on Disruptive, Innovative and Emerging Technologies in the Nuclear Industry*, Toronto, Canada.

Zubair, M., & Bibi, A. (2025). Enhancing safety of nuclear power plant by human error detection and identification through AI techniques. *Results in Engineering, 27*, Article 106616. https://doi.org/10.1016/j.rineng.2025.106616

Declaration

Consent for publication

Not Applicable

Availability of data and materials

The publisher has the right to make the data public

Conflict of Interest

The authors declared no conflict of interest

Ethical Considerations

Not applicable

Competing interest

The authors report no conflict or competing interest

Funding

The author declared no source of funding

Authors' Contribution

All components of the work were carried out by the author