

# Digital Twin Driven Hybrid LSTM And Isolation Forest Framework For Predictive Failure Detection And Autonomous Self Healing In Servers

Karthik. M<sup>1</sup>, Sanjay Jerene. M<sup>2</sup>, Ranjith. R<sup>3</sup>, Shafee Hamath. H<sup>4</sup>, Vigneshkumar. R<sup>5</sup>

<sup>1, 2, 3, 4, 5</sup> Dept of Computer Science and Engineering,

<sup>1, 2, 3, 4, 5</sup> Sri Manakula Vinayagar Engineering College, Puducherry, India

**Abstract-** Cloud server infrastructures form the foundation for technical services, where the services will be highly available, critical for business continuity and user satisfaction. Traditional systems notify the user after the failures had occurred. It is completely based on threshold-based alerts and results in performance-degradation, downtime, inefficient resource utilization and delayed recovery. To address this problem, our proposed framework twins the cloud server digitally to predict the anomalies and self-heal the cloud server environments. The framework continuously mirrors telemetry data such as CPU utilization, memory usage, disk I/O and network throughput into a virtual replica of the physical infrastructure. A hybrid anomaly detection model that combines Long Short-Term Memory (LSTM) networks and Isolation Forest is used to identify early-stage performance degradation. The proposed system results in reduced downtime, efficient resource utilization when compared to traditional monitoring techniques. The framework demonstrates the feasibility of autonomous and intelligent cloud infrastructure management.

**Keywords:** Digital twin, cloud computing, LSTM, Isolation Forest, anomaly detection, self-healing systems, predictive analytics, autonomous infrastructure

## I. INTRODUCTION

Cloud Computing contributes to various applications in education, artificial intelligence and so on. It operates on the basis of highly dynamic workload management, where system failures or performance degradation can lead to significant financial losses and poor user experience. As organisations increasingly depend on cloud-based services, maintaining availability and reliability has become a major challenge.

Traditional systems notify the user after the predefined thresholds have exceeded. Though they provide visibility to system performance, they are reactive in nature and need manual maintenance for recovery. This method is

not the best where telemetry data is generated continuously and system behaviour. This approach is not suitable for large-scale distributed infrastructures where telemetry data is generated continuously and system behavior changes rapidly.

Digital twin technology provides a new paradigm for infrastructure management by creating a continuously synchronized virtual representation of the physical system. When combined with machine learning techniques, the digital twin becomes capable of forecasting future system states, identifying anomalies before failure occurs, and evaluating corrective actions in a simulated environment.

This method proposes a hybrid predictive and self-healing framework that integrates digital twin with LSTM and Isolation Forest model. This enables early detection of performance degradation and autonomous execution of validated recovery strategies. The main aim of this study are:

- To predict the anomalies of the telemetry data for cloud servers.
- To implement simulation-driven self-healing using a digital twin.
- To reduce downtime through proactive management.

## Problem Statement

Cloud server generates large volumes of telemetry data that reflect system performance and health. Manual intervention of this data is difficult and leads to fault occurrence. Traditional systems notify the user after the predefined thresholds have exceeded.

Though predictive models have been introduced for anomaly detection, they are usually implemented as standalone components without integration into a closed-loop autonomous framework. Similarly, digital twin are used for visualization and simulation.

Another limitation is the absence of validation mechanisms for automated corrective actions. Self-healing strategies without prior evaluation can introduce operational risks and instability. Therefore, there is a need for an intelligent framework that can:

- predict failures at an early stage
- validate corrective actions in a virtual environment
- perform autonomous recovery

## II. LITERATURE REVIEW

The predictive analytics in cloud has received significant attention due to increasing demand for reliable resource management. Digital twin has emerged as a solution for monitoring real-time physical systems. Sheridan Gomes et al.(2022) developed an OpenStack digital twin to monitor and visualize system performance. They showed a better operational insight into the infrastructure. However, they did not consider predictive analytics and automated fault remediation.

Ahmed et al. (2020) developed a digital twin-driven method for energy optimization in data centers using simulation-based workload analysis, demonstrating improvements in resource utilization and power efficiency. Deepika Saxena (2025) developed a federated learning-based predictive maintenance model for distributed cloud environments. The model was developed to improve prediction accuracy while maintaining privacy during collaborative training. However, the computational overhead and complexity made it difficult to apply in real-time cloud scenarios.

LSTM models are used for capturing long-term temporal dependencies, while Isolation Forest is widely used for unsupervised detection of rare abnormal behavior in high-dimensional telemetry data. The combination of these techniques enables the identification of both gradual performance degradation and sudden system failures, which cannot be achieved using a single model.

The existing studies focus on individual aspects such as monitoring, prediction and energy optimization. The proposed method addresses this research gap by incorporating LSTM-Isolation Forest anomaly detection model within a real-time digital twin to enable proactive and automated recovery in cloud server.

## III. METHODOLOGY

The proposed system implements a **Digital Twin-driven predictive self-healing framework** for cloud server environments. The methodology consists of data acquisition, preprocessing, hybrid anomaly detection, and autonomous recovery.

### A. Data Collection

System-level telemetry data is collected to represent the operational state of cloud servers. The dataset includes key performance metrics such as CPU utilization, memory usage, disk activity, and network throughput. These parameters reflect system health and workload conditions.

The data is captured at regular time intervals, forming a **multivariate time-series dataset**. Continuous synchronization between the physical server and the digital twin ensures that the virtual model accurately mirrors real-time system behavior.

### B. Data Preprocessing

Raw telemetry data is transformed into a structured format suitable for machine learning. Missing values are handled using interpolation techniques to preserve temporal continuity.

Normalization is applied to ensure uniform scaling and improve model convergence. The data is then converted into sequences using a **sliding window approach**, enabling the model to capture temporal dependencies effectively.

### C. Hybrid Model Architecture

A hybrid anomaly detection model is used to improve detection accuracy by combining temporal learning and unsupervised anomaly detection.

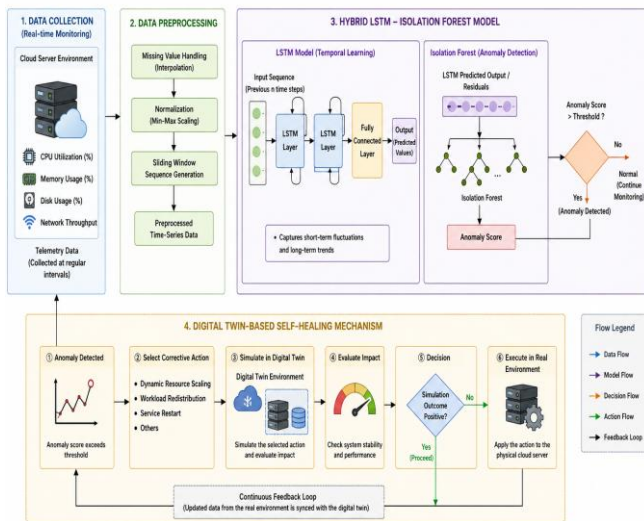


Fig.1. Architecture diagram for Digital-twin driven framework

1. LSTM Model

The Long Short-Term Memory (LSTM) network is used to analyze time-series data. It captures both short-term fluctuations and long-term trends in system performance. The model is trained using backpropagation through time with adaptive optimization techniques.

2. Isolation Forest

Isolation Forest is used for unsupervised anomaly detection. It identifies anomalies by isolating data points through random partitioning. Data points that are isolated quickly are considered anomalies.

3. Hybrid Integration

The outputs from LSTM and Isolation Forest are combined to generate an anomaly score. This hybrid approach enables detection of both **gradual performance degradation** and **sudden anomalies**.

D. Digital Twin-Based Self-Healing

When the anomaly score exceeds a predefined threshold, corrective actions are triggered. These actions include:

- Dynamic resource scaling
- Workload redistribution
- Service restart

Before applying any action to the physical system, it is first simulated in the digital twin environment. If the

simulation produces a positive outcome, the action is executed in the real system.

This closed-loop mechanism ensures **safe, reliable, and automated recovery**.

IV. EXPERIMENTAL RESULTS

The evaluation was conducted using time-series data from different workload conditions in the server environment. The LSTM model was trained to learn patterns, and Isolation forest was used to compute anomaly scores based on real-time deviations. Traditional systems notify the user after the predefined thresholds have exceeded.

But, the LSTM model identifies anomalies at an earlier stage by learning long-term temporal dependencies in the telemetry data. The hybrid LSTM-Isolation Forest model showed the best performance among all the approaches. By combining temporal prediction error with unsupervised anomaly scoring, the model successfully detected both performance degradation and abnormalities. This notifies the system before failure happens.

Model	Detection Accuracy	Downtime Reduction	Resource Utilization Efficiency
Threshold Monitoring	Low	Minimal	Moderate
LSTM	High	Moderate	High
Hybrid Model	Very High	Significant	Very High

Performance Comparison of Different Approaches

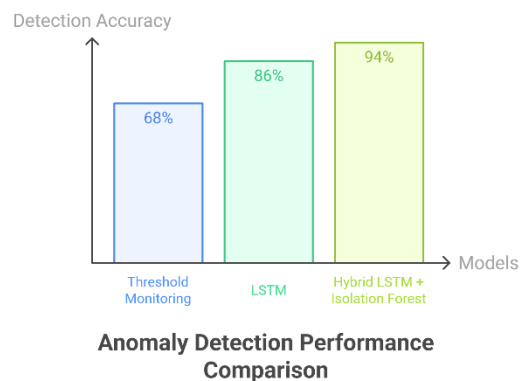
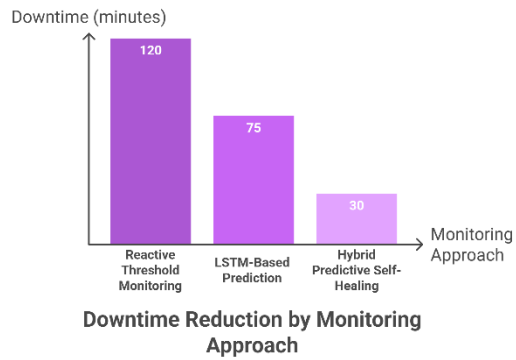


Fig. 2. Anomaly detection performance comparison



**Fig.3. Downtime reduction by monitoring approach**

The digital twin integration has to be checked before execution to reduce the risk of incorrect automated recovery and to improve the stability of the system.

## V. DISCUSSION

The results show that predictive models are more effective than traditional models. The hidden temporal patterns can be recognized by the LSTM model; however, using LSTM alone may not always detect rare and sudden anomalies. To overcome this limitation, the hybrid model with Isolation Forest is used. The LSTM model identifies anomalies at an earlier stage by learning long-term temporal dependencies in the telemetry data, and the Isolation Forest detects abnormal behavior. Predictive scaling allows resources to be allocated based on future demand rather than current overload conditions. This results in better workload distribution and reduced operational cost.

### A. Limitations

The prediction accuracy depends on the training data. Learning models require tuning of hyperparameters and sufficient training data to achieve optimal performance. The proposed framework currently operates on historical telemetry data for predictive analysis. Real-time streaming implementation may introduce additional challenges related to latency and scalability.

## VI. CONCLUSION

The study develops LSTM with unsupervised anomaly detection using Isolation Forest to identify performance degradation. The digital twin provides a simulation environment for validating corrective actions before applying them to the physical system.

The results demonstrate that the model improves anomaly detection capability and reduces system downtime.

Predictive scaling also enhances resource utilization and reduces the need for manual intervention. The proposed framework highlights the potential of integrating digital twin technology with machine learning to create autonomous and intelligent cloud infrastructures. Instead of reacting to failures after they occur, the system enables proactive management by forecasting system behavior and executing validated recovery strategies.

## VII. ACKNOWLEDGMENT

This is a great pleasure and immense satisfaction to express my deepest sense of gratitude and thanks to everyone who has directly or indirectly helped me in completing my paper work successfully. I am very thankful to acknowledge Dr. N. Danapaquaim (HOD, Computer Engineering). I present my sincere thanks to Dr. V.S.K. Venkatachalapathy, Principal for moral support and providing excellent infrastructure in carrying out the paper work. No words are sufficient to express my gratitude to my family for their unwavering encouragement. I also thank all friends for being a constant source of my support.

## REFERENCES

- [1] D. Saxena, A. K. Singh, and P. Gupta, "A Self-Healing and Fault-Tolerant Cloud-Based Digital Twin Processing Management Model (SF-DTM)," *IEEE Transactions on Industrial Informatics*, vol. 21, no. 3, pp. 2451–2462, 2025.
- [2] R. Ogunsakin, M. Adeyemi, and C. Li, "Towards Self-Adaptive Cloud Management with Digital Twins," *Journal of Cloud Computing*, vol. 12, no. 4, pp. 210–224, 2023.
- [3] M. Ahmed and Y. Lee, "Digital Twins for Data Centers: Energy Efficiency and Resource Optimization," *Sustainable Computing: Informatics and Systems*, vol. 31, 100647, 2021.
- [4] R. van Dinter, L. Peters, and H. Bosch, "Reference Architecture for Digital Twin-Based Predictive Maintenance Systems," *Computers & Industrial Engineering*, vol. 178, 109040, 2023.
- [5] R. van Dinter, M. de Vries, and P. Janssen, "Predictive Maintenance Using Digital Twins: A Systematic Literature Review," *Information and Software Technology*, vol. 146, 106908, 2022.
- [6] H. Li and P. Zhao, "Digital Twin-Driven Anomaly Detection for Cloud Infrastructure," *Future Generation Computer Systems*, vol. 138, pp. 32–45, 2023.
- [7] S. Gomes, A. N. Toosi, and B. Ens, "Digital Twin of a Cloud Data Centre: OpenStack Cluster Visualisation,"

- Springer Book Chapter in Cloud Computing for Data Analytics, pp. 115–130, 2022.
- [8] A. G. Wermann and J. A. Wickboldt, "KTWIN: A Serverless Kubernetes-Based Digital Twin Platform," *Computer Networks*, vol. 259, 111095, 2025.
- [9] R. Achary and C. J. Shelke, "Automated Cloud Scaling with GRU, LSTM, VARMAX Predictions," *2024 International Conference on Innovation and Novelty in Engineering and Technology (INNOVA)*, Vijayapura, India, 2024, pp. 1-5, doi: 10.1109/INNOVA63080.2024.10846983.
- [10] S. S. R. Selvaraj, V. Suresh, I. Gugan and Y. D, "A Hybrid Machine Learning Framework for Insider Threat Detection and Response," *2025 9th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2025, pp.1369–1374,doi:10.1109/ICECA66444.2025.11383492.
- [11] S. Kaur, P. Chowhan and A. Sharma, "A Novel Hybrid Deep Learning-Based Framework for Intelligent Anomaly Detection in Smart Meters," in *IEEE Access*, vol. 13, pp. 107022-107034, 2025, doi: 10.1109/ACCESS.2025.3581257