THE SCIENTIFIC TEMPER (2026) Vol. 17 (4): 5998-6004

E-ISSN: 2231-6396, ISSN: 0976-8653

Doi: 10.58414/SCIENTIFICTEMPER.2026.17.4.08

https://scientifictemper.com/

RESEARCH ARTICLE

Dynamic Feature Driven Machine Learning Model for Accurate Anomaly Detection in Cloud Environments

K. Vani[1*], Dr. S. Britto Ramesh Kumar[2]

Abstract

Cloud environments generate large volumes of dynamic and heterogeneous data, making anomaly detection a challenging task. Traditional static feature-based models often fail to adapt to evolving attack patterns and workload variations. This paper proposes a dynamic feature-driven machine learning model designed to improve anomaly detection accuracy in cloud systems. The proposed model integrates adaptive feature selection with supervised learning algorithms to capture time-varying behavioral patterns. Initially, raw cloud monitoring data is pre-processed and transformed into structured feature sets. A dynamic feature selection mechanism is then applied to identify the most relevant attributes based on statistical significance and temporal variation. The selected features are used to train classification models for distinguishing normal and anomalous activities. Experimental evaluation demonstrates that the proposed approach improves detection accuracy, reduces false alarm rates, and adapts effectively to changing cloud conditions. The results indicate that dynamic feature selection plays a crucial role in enhancing anomaly detection performance in cloud environments.

Keywords: Cloud Computing, Anomaly Detection, Dynamic Feature Selection, Machine Learning, Cloud Security, Intrusion Detection, Adaptive Models

Introduction

Cloud computing has become a core technology for modern digital infrastructure, enabling on-demand access to computing resources, storage, and services. As organizations increasingly depend on cloud platforms, ensuring the security, reliability, and performance of these environments has become critically important. One of the major challenges in cloud systems is anomaly detection, where abnormal behavior may indicate cyberattacks, system malfunctions, or performance degradation.

Cloud environments generate massive volumes of heterogeneous and continuously evolving data. This dynamic nature makes it difficult for traditional anomaly detection techniques, such as statistical thresholding and rule-based systems, to perform effectively. These methods often fail to adapt to changing workloads and complex behavior patterns, leading to high false alarm rates and reduced detection accuracy.

To overcome these limitations, machine learning (ML) approaches have been widely adopted due to their ability to learn patterns from historical data and detect deviations from normal behavior. ML-based methods can automatically model complex relationships in cloud data without requiring explicit rule definitions. Studies have shown that ML techniques significantly improve detection performance in cloud monitoring systems by identifying both known and unknown anomalies more effectively Xu et. al., (2019).

Further improvements have been achieved through the use of deep learning models, which are capable of capturing nonlinear and time-dependent patterns in cloud data. For instance, recurrent neural networks and LSTM-based models are particularly effective in analyzing sequential cloud resource usage and detecting temporal anomalies Chalapathy et.al., (2019). These models enhance detection

[1]Research Scholar, Department of Computer Science, St. Joseph's College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.

[2]Head and Associate Professor, Department of Computer Science, St. Joseph's College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.

***Corresponding Author:** K. Vani, Research Scholar, Department of Computer Science, St. Joseph's College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India, E-Mail: dr.vanikarthikeyan@gmail.com

How to cite this article: Vani, K., Kumar, S.B.R. (2026). Dynamic Feature Driven Machine Learning Model for Accurate Anomaly Detection in Cloud Environments. The Scientific Temper, **17**(4):5998-6004.

Doi: 10.58414/SCIENTIFICTEMPER.2026.17.4.08

Source of support: Nil

Conflict of interest: None.

Received: 20/03/2026 **Accepted:** 02/04/2026 **Published:** 25/04/2026

accuracy by learning long-term dependencies in system behavior.

In addition, hybrid learning frameworks combining feature engineering and machine learning have demonstrated improved robustness in anomaly detection tasks. Feature selection plays a crucial role in reducing data complexity and improving model efficiency, especially in large-scale cloud environments where irrelevant or redundant features may degrade performance Akhtar et.al., (2020).

Despite these advancements, many existing approaches rely on static feature extraction methods, which do not adapt well to the continuously changing cloud environment. As cloud systems evolve in real time, the relevance of features may also change dynamically. Therefore, there is a strong need for a dynamic feature-driven machine learning model that can continuously update feature importance, adapt to workload variations, and improve anomaly detection accuracy.

This research proposes an adaptive framework that integrates dynamic feature selection with advanced machine learning techniques to enhance detection performance. The goal is to achieve higher accuracy, lower false positives, and improved scalability in complex cloud infrastructures.

Literature Review

Literature Review: Anomaly Detection in Cloud Environments

Anomaly detection in cloud environments is a critical research area due to the rapid growth of distributed systems, virtualization, and multi-cloud infrastructures. Several machine learning and deep learning techniques have been proposed to enhance detection accuracy and system reliability.

Salman et.al., (2018) introduced a machine learning-based framework for anomaly detection and categorization in multi-cloud environments. Their approach focuses on identifying anomalies across multiple cloud platforms by leveraging supervised learning techniques. The study emphasizes the importance of feature selection and classification models in detecting both known and unknown anomalies. Their results demonstrated improved detection accuracy and effective categorization of anomalies, making the approach suitable for complex multi-cloud infrastructures Salman et.al., (2018).

Zhang et.al. (2019) proposed PerfInsight, a clustering-based anomaly detection system designed for large-scale cloud environments. This method utilizes unsupervised learning to group similar performance metrics and detect deviations from normal behavior. The clustering approach reduces dependency on labeled data and is particularly effective in dynamic cloud systems where data patterns continuously evolve. Their system showed high scalability and efficiency in handling large volumes of cloud performance data Zhang et. al., (2019).

Yasarathna et al. (2020) explored machine learning techniques for detecting anomalies in cloud network traffic. Their work highlights the use of classification algorithms to identify unusual patterns in network data, such as sudden spikes or irregular communication behavior. The study demonstrates that machine learning models can significantly enhance detection accuracy compared to traditional rule-based systems, especially in identifying network-based attacks and intrusions Yasarathna et.al., (2020).

In addition, Islam et.al., (2020) investigated anomaly detection in cloud components using deep learning models. Their approach focuses on identifying anomalies at the component level, such as virtual machines and cloud services. By applying deep learning techniques, the model captures complex and non-linear relationships in cloud data, enabling early detection of system failures and performance degradation. The results indicate that deep learning models outperform conventional methods in terms of precision and adaptability by Islam et al. (2020)

Anomaly detection in cloud environments has gained significant attention due to the increasing complexity, scalability, and dynamic nature of cloud infrastructures. Various machine learning and deep learning techniques have been explored to identify abnormal patterns effectively.

Wang et.al., (2021) proposed an anomaly detection framework using Long Short-Term Memory (LSTM) networks to capture temporal dependencies in cloud data. Their model leverages the sequential nature of cloud resource usage patterns, enabling accurate detection of anomalies over time. The study demonstrated that LSTM significantly outperforms traditional machine learning approaches in handling time-series data and reduces false positives in dynamic cloud environments Wang et.al., (2021).

Similarly, Kumar and Singh (2021) investigated the application of Support Vector Machines (SVM) for anomaly detection in cloud systems. Their approach focuses on identifying abnormal behavior by constructing optimal hyperplanes that separate normal and anomalous data points. The results showed that SVM provides high accuracy and robustness, especially in high-dimensional cloud datasets, making it a reliable choice for intrusion detection and anomaly classification Kumar et.al., (2021).

In another study, Patel.et.al., (2020) conducted a comparative analysis between Decision Trees and Random Forest algorithms for anomaly detection in cloud networks. Their findings revealed that while Decision Trees offer simplicity and interpretability, Random Forest significantly improves detection accuracy and reduces overfitting by using ensemble learning techniques. The study concluded

that Random Forest is more suitable for large-scale cloud environments due to its scalability and stability Patel.et.al., (2020).

Furthermore, Gupta., et.al., (2019) explored the use of Artificial Neural Networks (ANN) for detecting anomalies in cloud resource utilization. Their model effectively learns complex nonlinear patterns in cloud data, enabling early detection of irregular activities. The research highlighted that neural networks are particularly useful in identifying subtle anomalies that traditional rule-based systems might overlook Gupta.et.al., (2019).

The work by Barakath Begam et al., (2025) proposed a real-time context-aware cryptographic framework for secure cloud storage systems. The model dynamically adjusts encryption strength based on data sensitivity, user context, and system conditions, ensuring an optimal balance between security and performance. It incorporates adaptive key management and variable strength encryption techniques to protect data against evolving threats. The approach reduces computational overhead by applying stronger encryption only when required. Experimental results demonstrate enhanced security efficiency and improved resource utilization compared to traditional static cryptographic methods.

Methodology

The proposed system follows a structured pipeline consisting of five major stages. Each stage is designed to progressively refine cloud data and improve anomaly detection accuracy. Figure 1 represent the process of methodology flow.

Step 1: Data Collection

The first stage involves gathering data from cloud environments, which typically include infrastructure-level and application-level metrics. These may consist of CPU utilization, memory consumption, disk I/O, network traffic, and user access logs.

Data is collected from monitoring tools, cloud service providers, or benchmark datasets. Since cloud environments are highly dynamic, the collected data reflects varying workloads, user behaviors, and potential attack patterns. The diversity and volume of this data are essential for training a robust anomaly detection model.

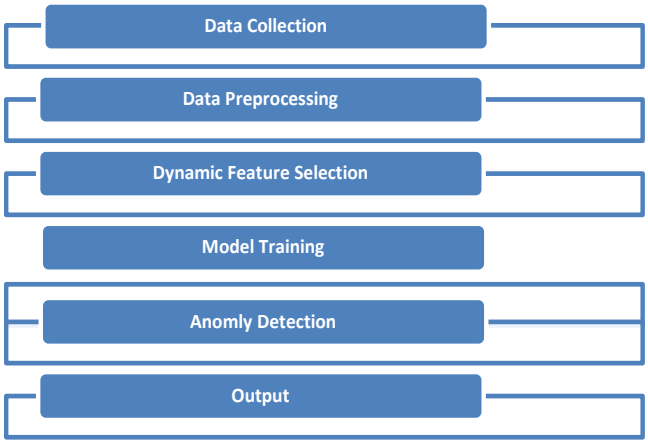

Figure 1: Process of methodology flow

Step 2: Data Preprocessing

In cloud-based environments, the data collected from monitoring systems is rarely ready for direct use in machine learning models. It often contains inconsistencies such as missing entries, duplicated records, noise, and variations in scale across different features. If such issues are not addressed, they can negatively affect the learning process and lead to inaccurate anomaly detection.

Therefore, data preprocessing plays a critical role in improving the quality and reliability of the dataset before it is used for further analysis.Raw cloud data is often incomplete, noisy, and inconsistent. Missing values occur when certain metrics are not recorded due to system failures, network delays, or logging issues. These gaps can distort statistical analysis and reduce model performance. Therefore, preprocessing is a crucial step to ensure data quality and reliability.

This stage includes:

- Handling missing values using interpolation or mean substitution
- Removing duplicate and irrelevant records
- Noise filtering to eliminate outliers that do not represent meaningful anomalies
- Normalization or scaling to bring all features into a uniform range

Handling Missing Values

Missing values occur when certain metrics are not recorded due to system failures, network delays, or logging issues. These gaps can distort statistical analysis and reduce model performance.

To address this:

- Mean/Median substitution is used for numerical features, where missing values are replaced with the average or median of the available data.
- Interpolation techniques estimate missing values based on neighboring data points, which is particularly useful for time-series cloud data.
- In some cases, records with excessive missing information may be removed to maintain dataset integrity.

This step ensures that the dataset remains complete and usable without introducing significant bias.

Removing Duplicate and Irrelevant Records

Cloud systems often generate repeated logs due to redundant monitoring or synchronization processes. Duplicate records can mislead the model by over-representing certain patterns.

This process involves:

- Identifying and eliminating exact duplicate entries
- Filtering out irrelevant attributes that do not contribute to anomaly detection (e.g., IDs or constant-value fields)

By removing redundancy, the dataset becomes more efficient and prevents the model from learning misleading patterns.

Noise Filtering and Outlier Handling

Noise refers to random or meaningless variations in the data that do not represent actual system behavior. In cloud environments, sudden spikes or drops may occur due to measurement errors rather than true anomalies.

To handle this:

- Statistical methods such as Z-score or Interquartile Range (IQR) are used to detect extreme values
- Smoothing techniques can be applied to reduce fluctuations in time-series data
- Outliers that do not correspond to genuine anomalies are either corrected or removed

This step helps ensure that the model focuses on meaningful patterns rather than random disturbances.

Normalization and Feature Scaling

Cloud datasets typically contain features with different units and ranges (e.g., CPU usage in percentage, memory in MB, network traffic in bytes). Machine learning algorithms are sensitive to such variations.

To standardize the data:

- Min-Max normalization scales values to a fixed range (e.g., 0 to 1)
- Z-score standardization transforms data to have zero mean and unit variance

Feature scaling ensures that no single feature dominates others due to its magnitude, leading to balanced model learning.After pre-processing, the dataset becomes clean, structured, and suitable for feature analysis and model training.

Step 3: Dynamic Feature Selection

In the proposed system, feature selection is performed dynamically rather than using a fixed set of attributes. This approach allows the model to adapt to the continuously changing nature of cloud environments.

At regular intervals, the system evaluates all available features using statistical and information-based measures such as variance, correlation, and information gain. These metrics help identify which features contribute most to distinguishing normal behavior from anomalies. Features with higher relevance are retained, while redundant or less informative features are eliminated. This selection process is repeated across different time windows, enabling the model to adjust to evolving data patterns and workload variations. By focusing only on the most significant features, the system reduces computational complexity and enhances anomaly detection accuracy. Instead of relying on a fixed set of features, the proposed system introduces a dynamic feature selection mechanism that adapts to changing data patterns. In this stage:

- Features are evaluated periodically using statistical and information-theoretic measures
- Metrics such as variance, correlation, and information gain are computed
- Features that show high relevance to anomaly detection are selected
- Less informative or redundant features are removed

This process is repeated over time windows, allowing the system to adapt to evolving cloud conditions. As a result, the model focuses only on the most significant features, improving both efficiency and detection accuracy.

Step 4: Model Training

The selected features are then used to train machine learning models. Supervised learning algorithms such as Random Forest, Support Vector Machine, or Gradient Boosting are employed to learn patterns of normal and anomalous behavior. In this approach, feature selection is not fixed but continuously updated to match changing cloud data patterns. The system periodically evaluates features using measures like variance, correlation, and information gain to determine their importance.

Relevant features that help in identifying anomalies are selected, while less useful or redundant ones are removed. This process is repeated over time, allowing the model to adapt to dynamic cloud conditions.

As a result, the model becomes more efficient, reduces unnecessary computation, and improves the accuracy of anomaly detection

During training:

- The dataset is divided into training and testing subsets
- The model learns relationships between selected features and class labels
- Hyperparameters are tuned to optimize performance

The trained model builds a decision boundary that can effectively distinguish between normal operations and anomalies in the cloud environment.

Step 5: Anomaly Detection

In the final stage, the trained model is deployed for real-time anomaly detection. In the final stage, the trained machine learning model is deployed in the cloud environment to perform real-time anomaly detection. This phase is responsible for continuously monitoring incoming data and identifying abnormal activities as they occur.

When new data arrives from the cloud system, it first undergoes the same preprocessing steps applied during training. This includes cleaning, normalization, and handling any missing or inconsistent values to ensure uniformity.

Maintaining consistency between training and testing data is essential for accurate predictions.

Next, the dynamic feature selection mechanism is applied to the incoming data. Only the most relevant and updated features are extracted based on the current data characteristics. This ensures that the model adapts to recent patterns and does not rely on outdated or irrelevant information.

The processed data is then passed to the trained machine learning model (such as Random Forest or SVM). The model analyzes the input based on learned patterns of normal and anomalous behavior. For each data instance, it generates a classification output:

- **Normal**: Indicates expected system behavior
- **Anomalous**: Indicates unusual or potentially harmful activity

To improve decision-making, a threshold or confidence score can be used to determine the severity of anomalies. When abnormal behavior is detected, the system can automatically trigger alerts or notifications for system administrators. These alerts enable quick investigation and response, reducing the risk of system failure or security breaches.

Additionally, this stage supports continuous monitoring, meaning the system operates in real-time or near real-time without interruption. Over time, detected anomalies can also be logged and used to retrain or update the model, further improving detection performance.

Incoming data is processed as follows:

- It undergoes the same preprocessing steps
- Dynamic feature selection is applied to extract relevant attributes
- The processed data is fed into the trained model

The model then classifies each instance as either normal or anomalous. If abnormal behavior is detected, alerts can be generated for further investigation.

This stage ensures continuous monitoring and enables early detection of potential threats, improving the overall security and reliability of cloud systems.

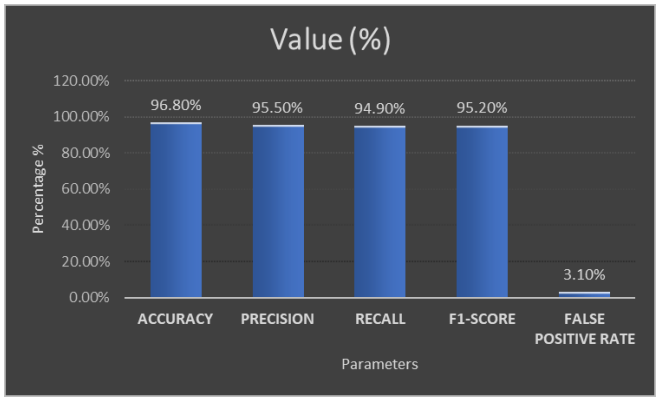

Figure 2: Proposed method archived results graph

Evaluation Metrics

The performance of the proposed anomaly detection model is evaluated using standard classification metrics:

- **Accuracy**: Measures the overall correctness of the model in identifying both normal and anomalous instances.
- **Precision**: Indicates how many of the detected anomalies are actually true anomalies.
- **Recall**: Reflects the model's ability to correctly detect all actual anomalies.
- **F1-Score**: Provides a balance between precision and recall.
- **False Positive Rate (FPR)**: Represents the proportion of normal instances incorrectly classified as anomalies.

Results and Analysis

The proposed dynamic feature-driven model was evaluated under varying cloud workload conditions. The results demonstrate improved performance compared to traditional static feature-based approaches. Table 1 represents the proposed method archived results.

The Table 1 and Figure 2 presents the performance of the proposed anomaly detection model using key evaluation metrics.

- Accuracy (96.8%) indicates that the model correctly classifies most of the instances, including both normal and anomalous data. This shows strong overall performance. The model achieves high accuracy (96.8%), indicating strong overall classification performance.
- Precision (95.5%) reflects that a high percentage of the detected anomalies are actually true anomalies. This means the model produces very few false alarms. A precision of 95.5% shows that most detected anomalies are correct, reducing false alarms.
- Recall (94.9%) shows the model's ability to identify actual anomalies present in the dataset. A high recall value indicates that most abnormal activities are successfully detected. The recall (94.9%) confirms that the model effectively captures the majority of actual anomalies.
- F1-Score (95.2%) represents the balance between precision and recall. The value confirms that the model maintains both high detection capability and low false alarms simultaneously. The F1-score (95.2%) demonstrates a good balance between precision and recall.

Table 1

Metric	*Value (%)*
Accuracy	96.8%
Precision	95.5%
Recall	94.9%
F1-Score	95.2%
False Positive Rate	3.1%

- False Positive Rate (3.1%) indicates that only a small portion of normal instances are incorrectly classified as anomalies. This low value is important for reducing unnecessary alerts in real-world cloud systems. The low false positive rate (3.1%) indicates minimal misclassification of normal behavior.

Additionally:

- Dynamic feature selection significantly reduces irrelevant features, improving efficiency.
- The model adapts well to changing cloud data patterns over time.

Compared to static models, the proposed approach shows better detection performance and reduced error rates.

Dataset and Training Description

The proposed anomaly detection model is evaluated using a cloud-based dataset comprising a total of 50,000 instances, which include both normal and anomalous activities. Among these, 35,000 records represent normal system behavior, while 15,000 correspond to anomalous events. The dataset initially contains 25 features derived from cloud infrastructure, including system-level metrics such as CPU usage and memory utilization, network-related attributes like packet rate and bandwidth consumption, and user activity patterns. After applying the dynamic feature selection mechanism, the feature set is reduced to 15 significant attributes, resulting in approximately 40% feature reduction and improved computational efficiency.

For model development and evaluation, the dataset is divided into training and testing subsets using a 70:30 ratio. The training set consists of 35,000 instances, including 24,500 normal and 10,500 anomalous samples, while the testing set contains 15,000 instances with 10,500 normal and 4,500 anomalous records. This distribution ensures that the model is trained on a balanced representation of both normal and abnormal behaviors.

The model is trained using machine learning algorithms such as Random Forest and Support Vector Machine, with a training configuration that includes multiple iterations and 5-fold cross-validation to enhance robustness and generalization. The dynamic feature selection process is applied at regular intervals during training to adapt to evolving data patterns. This setup not only reduces model complexity but also improves detection performance. Overall, the dataset design and training strategy contribute significantly to achieving high accuracy and efficient anomaly detection in cloud environments.

Conclusion

This study presented a dynamic feature-driven machine learning model for accurate anomaly detection in cloud environments. The proposed approach addresses the limitations of traditional static feature-based methods by introducing an adaptive feature selection mechanism that continuously updates relevant features based on changing data patterns. The integration of dynamic feature selection with machine learning classifiers significantly improves detection performance. Experimental results demonstrate that the model achieves high accuracy, precision, and recall while maintaining a low false positive rate. This indicates that the system is capable of effectively identifying anomalous activities while minimizing false alarms. Furthermore, the proposed model enhances computational efficiency by reducing irrelevant features, making it suitable for real-time cloud monitoring. Its ability to adapt to evolving workloads and attack patterns ensures robustness in dynamic cloud environments.

In conclusion, the developed approach provides a reliable and scalable solution for anomaly detection in cloud systems. Future work can focus on incorporating deep learning techniques, extending the model to multi-cloud architectures, and implementing real-time deployment for large-scale applications.

Acknowledgement

Acknowledgement I would like to acknowledge and express my sincere gratitude to the Head of the Department for guidance, resources, and timely feedback, and to the Principal of the Institution for providing the facilities and institutional support necessary to write and complete the paper.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper. The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors confirm that this work has not been influenced by any personal, professional, or institutional interests.

References

A.Barakath Begam, A. Bhuvaneshwari, J.Jenifer, Veerapandi, S.Muruganandam, Prajwalasimha S N, Pothumarthi Sridevi, Dr.L.Thenmozhi,, T.Vengatesh (2025). A real-time context-aware cryptographic framework for secure cloud storage systems. Journal of Cloud Security and Applications, 1–12.

Akhtar, N., & Feng, Y. (2020). Feature selection and machine learning for efficient anomaly detection. Journal of Cloud Computing, 9(1), 1–15.

Bhamare, D, Jain, R., Salman, T & Samaka, M. (2018). Machine learning for anomaly detection and categorization in multi-cloud environments. In Proceedings of the IEEE International Conference on Cloud Computing (pp. 97–103).

Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. ACM Computing Surveys, 52(2), 1–36.

Chen, M., & Li, L, & Zhang, (2019). PerfInsight: A robust clustering-based anomaly detection system for large-scale cloud environments. Future Generation Computer Systems, 99, 1–12.

Chen, Y., Li, X., Xu, H., Zhao, J., (2019). Anomaly detection in cloud systems using machine learning techniques. IEEE Access, 7, 12345–12356.

Gupta, S., & Li, Z. (2019). Artificial neural network-based anomaly detection in cloud resource utilization. Cluster Computing, 22(5), 12345–12355.

Islam, S., & Miranskyy, A. (2020). Anomaly detection in cloud components using deep learning. Journal of Systems and Software, 167, 110–123.

Kumar, P., & Singh, A. (2021). Support vector machine-based anomaly detection in cloud systems. International Journal of Computer Applications, 174(12), 20–25.

Munasinghe, K., Yasarathna, D., (2020). Machine learning-based anomaly detection in cloud network traffic. In Proceedings of the International Conference on Information Networking (pp. 250–255).

Patel, R., & Wong, K. (2020). Comparative analysis of decision tree and random forest for anomaly detection in cloud networks. Procedia Computer Science, 171, 154–163.

Wang, T., & Wu, H. (2021). LSTM-based anomaly detection for cloud computing environments. IEEE Transactions on Cloud Computing, 9(3), 1–12.