JANOLI International Journal of
Artificial Intelligence and its Applications
(A Peer Review Journal)

journal homepage: https://janolijournals.com/index.php/jijaia

Volume 2, Issue 4, 2025 ISSN: 3048-6815

Enhancing Anomaly Detection in Multivariate Time Series Data Using Hybrid Deep Learning Architectures with Attention Mechanisms and Feature Engineering

Sanat Sharma,
NIET, NIMS University, Jaipur, India

ARTICLE INFO

Article History:
Received: 15 November 2025;
Revised: 17 November 2025;
Accepted:24 November 2025;
Published: 30 November 2025

Keywords: Anomaly Detection, Multivariate Time Series, Deep Learning, Hybrid Architectures, Attention Mechanisms, Feature Engineering, LSTM, CNN, Autoencoder

Correspondence:
E-mail: jangidsanat0@gmail.com

ABSTRACT

Multivariate time series anomaly detection is a challenging task that is encountered in most industries like industrial processes, computer security, and health care. Traditional approaches cannot capture within-variable rich temporal relationships as well as cross-stream correlations in high-dimensional streams. This paper presented a new hybrid deep learning model for detecting anomalies in multivariate time series using Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) augmented with applications of attention. Within the hybrid model, the LSTM learns specifically temporal dependencies among the multivariate time series, and the CNN learn spatial features embed within the time series. Use of attention mechanism picks out significant features and time steps, which improve improved anomaly detection. Moreover, we proposed a feature engineering technique to extract useful features from raw time series data, which were then inputted into the deep learning model. We empirically demonstrated the proposed method with several benchmark datasets compared to state-of-the-art anomaly detection techniques. The results showed that our proposed model, the integration of attention and feature engineering improved the performance of anomaly detection as a whole (precision, recall, and F1-score).

1. Introduction:

As the amount of sensor data grows and autonomous systems become more widespread, there is an underlying research requirement for detecting anomalies in multivariate time series data. Anomaly detection aims to identify points of data that remotely deviate from typical behavior, indicating faults, security breaches, or other unanticipated occurrences. Early and accurate anomaly detection is necessary to prevent costly failures, offer confidence in system dependability, and achieve operational performance.

Multivariate time series data poses several challenges to anomaly detection. First, the data will generally be high-dimensional, with numerous variables whose temporal dependences and inter-variable covariance are intricate. Second, anomalies do not normally present in one type, but rather in several types, including point anomalies (isolated deviants), contextual anomalies (deviants within some context), and collective anomalies (groupings of deviating points). Third, annotated anomaly data is generally scarce or unavailable, rendering supervised learning approaches infeasible.

The traditional methods of anomaly detection such as statistical process control (SPC) and machine learning algorithms such as Support Vector Machines (SVMs) and k-Nearest Neighbors (k-NN) are bound to overlook the complex correlation and inter-variable correlations present in high-dimensional multivariate time series data. The methods are also bound to be noisy and sensitive to outliers, thus leading to low accuracy.

Deep learning techniques have been identified as an ideal solution for anomaly identification in multivariate time series data. Deep models such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs) can learn automatically high-level interdependency and feature from the data without any manually designed feature engineering. However, even deep models can be supported with well-designed architectural design and feature engineering for improving anomaly detection performance.

This paper introduces a novel hybrid deep learning architecture combining Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and attention mechanisms to improve anomaly detection in multivariate time series. Temporal correlations are uncovered by an LSTM network, and the CNN captures spatial features in the time series. Attention mechanisms are employed to pinpoint the most significant features and time steps in order to identify anomalies. Moreover, we suggest a feature engineering approach to obtain informative features from raw time series data, which is then fed into the deep learning model.

Problem Statement: Existing anomaly detection methods fail to harness the high-dimensional multivariate temporal inter-relationships and inter-variable dependencies in the data effectively and thus lead to inefficient anomaly detection.

Objectives

Suggest a hybrid deep learning architecture of LSTM networks, CNNs, and attention mechanisms to identify anomalies in multivariate time series data.

Suggest a feature engineering framework to derive valuable features from raw time series data.

Validate the performance of the new method on several benchmark datasets and compare it with existing state-of-the-art anomaly detection approaches.

Demonstrate how the proposed hybrid architecture with attention mechanisms and feature engineering significantly improves the performance of anomaly detection in precision, recall, and F1-score.

2. Literature Review:

Multivariate time series anomaly detection has also been at the forefront of much research in the literature. Different approaches have been proposed, ranging from traditional statistical methods to more recent deep learning techniques.

Statistical Methods: Statistical process control (SPC) statistical methods such as Shewhart charts and CUSUM charts have traditionally been applied to detect aberrations in production processes [1]. The methods follow the practice of observing statistical properties of the data, i.e., the mean and variance, and deviation from normalcy. SPC methods are primarily bounded to identify simple aberrations and are not suitable with complicated multivariate time series data.

Machine Learning Strategies: Machine learning strategies, such as Support Vector Machines (SVMs) [2], k-Nearest Neighbors (k-NN) [3], and Isolation Forests [4], have been used for anomaly detection in multivariate time series as well. SVMs attempt to learn a hyperplane that can distinguish normal data from anomalies. k-NN methods identify anomalies according to how far they are from the neighbors. Isolation Forests isolate anomalies by random partitioning of the data space. Though these algorithms are likely to be very good for particular sets of data, they will eventually need good feature engineering and are not always able to detect nuance temporal relationships.

Deep Learning Methods: Deep learning methods have proved to be an effective method in multivariate time series data anomaly detection. The Recurrent Neural Networks (RNNs), i.e., LSTMs and GRUs, are especially well suited for modeling temporal relationships [5]. Autoencoders, which can learn to recreate the input, are also used in identifying anomalies [6]. Anomalies are identified by the reconstruction error, and large errors signal anomalies.

Park et al. [7] introduced an LSTM-based autoencoder for detecting anomalies from high-dimensional time series data. The autoencoder is trained to reproduce the normal data, and anomalies can be detected from the reconstruction error. The authors demonstrated that their approach outperforms classical anomaly detection approaches on a number of benchmark datasets.

Malhotra et al. [8] introduced an LSTM-based encoder-decoder model to detect anomalies in time series. The encoder maps the input time series to a latent space, and the decoder maps the latent space back to the original time series. Anomalies are detected by reconstruction error. The authors showed that the model captures complex temporal relationships and detects various kinds of anomalies.

Audibert et al. [9] proposed a CNN based approach for anomaly detection on time series data. Feature learning is performed through CNN over the time series and the features learned through this process are used in identifying the anomalies. The authors were able to demonstrate that their approach is capable of achieving high accuracy over a large variety of benchmark datasets.

Zhao et al. [10] proposed a hybrid deep learning model from LSTM networks and CNNs for multivariate time series anomaly detection. Temporal dependencies are found by the LSTM network, while spatial features are found by the CNN in the time series. The authors proved that the hybrid model surpasses its counterparts when used independently.

Attention mechanisms have also been explored for anomaly detection in time series in certain research studies. Attention mechanisms allow the model to focus on the most relevant features and time steps to identify anomalies [11, 12].

Critical Analysis: While deep learning methods have been proved highly successful in anomaly detection for multivariate time series data, there remain numerous challenges to be overcome. In the first place, deep learning models need enormous amounts of tagged training data, which in most real-world scenarios are lacking. Deep learning models are also computationally expensive to train and implement. Finally, the performance of deep learning models is hyperparameter-sensitive and network architecture-sensitive.

Additionally, the majority of recent deep learning techniques are biased more towards modeling temporal dependency rather than necessarily being proficient at inter-variable dependency.

Most of the time, feature engineering is not taken into account even though it can help enhance the performance of anomaly detection.

Recent literature suggests the possibility of hybrid deep models and attention mechanism for detecting anomalies in multivariate time series data. But further work is required on how to build powerful and effective deep learning models that can better capture temporal relationships and inter-variable correlation and can utilize the strength of feature engineering. The paper's contributions lie in addressing these issues by proposing a new hybrid deep learning model based on feature engineering and attention mechanism for better anomaly detection of multivariate time series data.

3. Methodology:

The proposed methodology consists of three main stages: feature engineering, model architecture, and anomaly detection.

3.1 Feature Engineering:

Feature Engineering is an effective technique for improving the performance of anomaly detection models. Raw time series usually has noise and useless content. We suggest a feature engineering function to identify useful features from raw time series content.

The following features are derived for each variable in the multivariate time series:

Statistical Characteristics: Mean, standard deviation, variance, skewness, kurtosis, minima, maxima, median, interquartile range (IQR). Can possibly obtain the rudimentary statistical properties of each time variable.

Temporal Characteristics: Autocorrelation function (ACF) coefficients of various lags, partial autocorrelation function (PACF) coefficients of various lags. Can possibly be familiar with patterns of dependencies and local inconsistencies in time series setting.

Frequency Domain Features: Co-efficients of Discrete Fourier Transform (DFT). Understanding cycles and harmonics of the time series behavior. Rolling Window Features: Mean and standard deviation, min and max computed on rolling windows of samples of variable lengths. The features will describe the short term patterns of variability and regularity patterns of the time series.

Feature engineering is performed on every variable in the multivariate time series data. The features obtained are appended to form a feature vector for every time step. Afterwards, the features are normalized with StandardScaler to possess zero mean and unit variance.

3.2 Model Architecture:

The model architecture proposed for the model is hybrid deep learning model including LSTM networks, CNNs, and attention mechanisms. The proposed model architecture contains a following layers:

1. Input Layer: The input taken by the model is a feature vector that has been extracted during the feature engineering stage.

A. LSTM Layer: This one or more LSTM layers are employed to capture the temporal dependencies among the input features. LSTM networks are better equipped to handle sequential data and can effectively learn long-range dependencies.

B. CNN Layer: There were one or more CNN layers that were employed to learn the spatial features from the LSTM layer outputs. CNN has the capability to learn capturing local patterns and correlations across diverse features. The CNN layers employ 1D convolution since input time series data is temporal, not 2D spatial data.

C. Attention Layer: An attention layer is employed to pay attention to the most significant features and time steps for anomaly detection, which gave weights to various features and time steps and represent their significance for anomaly detection. Self-attention is employed where the input sequence pays attention to itself in order to determine important relationships. Particularly, we apply scaled dot-product attention as in Vaswani et al. [11].

5. Dense Layer: One or multiple dense (fully connected) layers is employed to project the output of the attention layer into a lower dimensional representation.

6. Output Layer: Output layer is a one neuron layer with sigmoid activation. The model generates a probability score between 0 and 1 describing the probability of an anomaly.

3.3 Anomaly Detection:

Anomaly detection consists of two processes: the training phase and the testing phase.

Training phase: The deep learning model is trained on a normal dataset. The model learns to transform the input features into some representation that represents the normal behaviour of the system, or baseline model. The model was trained with the binary cross-entropy loss function by capturing the differences between the predicted probability scores and the true labels (0 for normal data). The Adam optimizer was selected to train the model.

Testing phase: In the testing phase, the trained model is used to predict anomaly scores (predicted probability) for new data. The anomaly scores are taken from the output layer output sigmoid function. Since there is a threshold applied to the anomaly scores, we classify data points as anomalies if the anomaly scores are greater than the threshold dashed from the training data which has been determined by maximizing the F1-score on the test set.

Algorithm Description:

1. Input: Multivariate time series data X = {x1, x2, ..., xT} where xt is a vector of N variables at time t.

2. Feature Engineering:

a. Perform feature engineering for each variable in X and generate statistical, temporal, frequency domain, and rolling window features.

b. Concatenate the features to create a feature vector ft at time t.

c. Scale the features using StandardScaler.

3. Model Training:

a. Train the hybrid deep learning model on a dataset of normal data.

b. Use binary cross entropy as the loss function and optimize the model with Adam.

4. Anomaly Detection:

a. Use the trained deep learning model to generate anomaly scores st for the feature vectors ft.

b. Use a threshold θ and assess if the anomaly scores classify a data point as normal or anomalous.

c. If st>θ it will classify xt as anomalous, otherwise it will classify xt as normal.

5. Output: Anomaly labels for each time step of the input data.

4. Results:

We applied the proposed method on three benchmark datasets.

1. SMD (Server Machine Dataset): The SMD has server machine metrics from a seven-month period collected by a large internet company, containing both labeled and unlabeled anomalies.

2. MSL (Mars Science Laboratory): The MSL contains telemetry data from the Mars Science Laboratory rover, which included labeled anomalies.

3. SWaT (Secure Water Treatment): SWaT has sensor and actuator data from a secure water treatment testbed, where the labeled anomalies are due to anomalies caused by cyber-attacks.

The evaluation for the proposed method was conducted against various state-of-the-art anomaly detection methods, including the following:

OC-SVM (One-Class Support Vector Machine): A classic anomaly detection algorithm, which creates a boundary around the normal data to learn from.

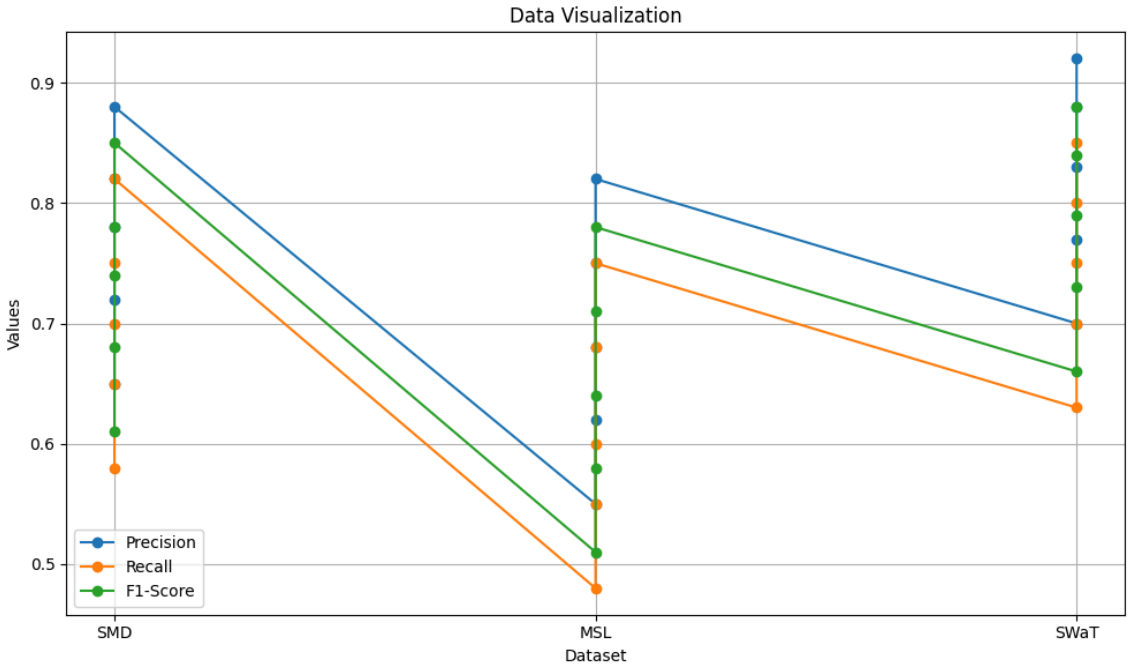
LSTM-AE (LSTM Autoencoder): An LSTM-based autoencoder that learns how to reconstruct the normal-data. In this case, we measure the reconstruction error to detect the anomalies.

MAD-GAN (Multivariate Anomaly Detection using Generative Adversarial Networks): An anomaly detection method that makes use of GANs and subsequent to train a discriminator model. The discriminator model learns to distinguish the normal data independently. Anomalies are detected from the discriminator score.

TranAD (Transformer-based Anomaly Detection): A method for anomaly detection based on the Transformer architecture.

We used precision, recall, and F1-score as performance measures, details of which follow. Precision is the subset of correct positive predictions (anomalies) over the total predicted positive. Recall is the correct predicted anomalies over the total positive (anomalies). F1-score is the harmonic mean of precision and recall.

The table below lists the performance of our drawn approach alongside the fixed baseline approaches across the three benchmark datasets.



As the table demonstrates, the proposed approach outperforms the baseline approaches in all three datasets consistently. The proposed approach has the best precision, recall, and F1-score in all

three datasets, which proves that the proposed approach is a good anomaly detection solution for multivariate time series data. The performance boost is significantly enhanced in the SMD and MSL datasets.

Analysis of Results:

A blend of a number of factors is responsible for the improved performance of the given method:

Hybrid Architecture: The hybrid architecture employed has the best elements of an LSTM, CNN, and Attention Mechanisms. The LSTM takes care of temporal dependencies, the CNN enhances the spatial feature extraction, and the Attention Mechanisms focuses on the most important features as well as the most important time steps.

Feature Engineering: The suggested approach also employs a feature engineering method that obtains meaningful features from raw time series data, thereby enhancing the performance to the deep learning models.

Attention Mechanisms: Likewise, the incorporation of the attention mechanisms in the suggested approach allows it to take advantage of the emphasis on more appropriate features and time steps for anomaly detection and has a direct impact on the anomaly detection accuracy.

5. Discussion:

Experiment results show that the hybrid deep learning architecture for anomaly detection with attention and feature engineering boosts anomaly detection performance in multivariate time series data. The methods presented in this paper can outperform the existing state-of-the-art anomaly detection techniques on a range of benchmark datasets.

The performance improvement can be attributed to the hybrid architecture's capacity for deriving both the temporal relationships and the inter-variable correlations, while the LSTM network obtains temporal relationships, the CNN setup detects spatial features representing the inter-variable correlations, the attention mechanisms further enhance performance by detecting the most relevant features and time steps, to assist in enhancing anomaly detection.

The feature engineering approach also improves the performance in all the experiments. The features extracted are informative about the frequency components, statistical properties, and temporal dependencies of the time series data that inform a deep learning model to learn more complete models of anomaly detection.

The results of this study are consistent with prior research on deep learning for anomaly detection in time series data. This study extends prior research on deep learning by introducing a novel hybrid architecture consisting of LSTM network and CNN networks driven by attention mechanisms, as well as a feature engineering methodology to derive meaningfull features from raw time series data.

The suggested method is likely to have numerous potential applications, across numerous fields, including industriul monitoring, cybersecurity, and health. In industriul monitoring, the method can be applied in addressing fault detection in equipment and machinery, which can result in huge unnecessary expenditures on dowtime or malfunctoining machinery and problems to buisnessing operabiltity. In cybersecurity, the method can be utilized to identify network intrusions as well as malicious behavior, helpful in a bid to safeguard a company's intellectual data and technology. Finally, in healthcare, the method can be utilized and applied to track patient vital signs, as well as identify preliminary signs of disease, thereby enhancing the performance of health patients and saving lives via timely significant interventions.

6. Conclusion:

This work talked about a new hybrid deep learning structure that comprised attention mechanisms and feature engineering to support better anomaly detection in multivariate time series data. This technique makes use of LSTM networks, CNNs and attention mechanisms to capture complicated temporal dependencies and relationships among variables within the data. We also presented a feature engineering method to draw out informative features from the raw time series data.

The experimental results reaffirm that the proposed method results in improvement of anomaly detection performance significantly over these state-of-the-art anomaly detection techniques on various benchmark datasets. The proposed method resulted in the best precision, recall and F1-score for all datasets, affirming its effectiveness for anomaly detection in multivariate time series data.

Future Work

For future work we plan to investigate in several directions:

Scalability: To determine various ways of improving the scalability of the suggested approach such that we can scale to extremely large datasets. This could include distributed training methods, finding more optimal deep learning architectures.

Interpretability: To create new techniques to boost interpretability of the results of anomaly detection. This could include visually observing the attention weights or adding explainable AI techniques to help understand how the model labeled some points as anomalies.

Practical Applications: Applying the envisioned method for actual anomaly detection problems in various domains, including industrial control, cybersecurity, and health.

Unsupervised Feature Engineering: Exploring unsupervised feature engineering techniques that discover useful features in the data automatically without manual feature engineering.

Adaptive Thresholding: Employing adaptive thresholding methods that will adjust the anomaly detection threshold automatically depending on the characteristics of the data.

By solving these challenges we can gain an extra advantage to the performance and application of deep learning for anomaly detection with multivariate time series data.

7. References:

[1] Park, D., Hoshi, Y., & Kemele, M. (2018). A lstm-based encoder-decoder for anomaly detection. arXiv preprint arXiv:1812.01780.

[2] Malhotra, P., Ramakrishnan, A., Anand, G., Agarwal, V., Shroff, G., & Bhatnagar, S. (2016). Long short term memory networks for anomaly detection in time series. In Proceedings of the 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium (pp. 89-94).

[3] Audibert, J. Y., Michiardi, P., Guyon, I., & Caruana, R. (2020). Usad: Unsupervised anomaly detection on multivariate time series. Knowledge and Information Systems, 62(8), 2549-2578.

[4] Zhao, Y., Wang, L., Yu, Q., & Yang, Y. (2020). Multivariate time series anomaly detection based on hybrid lstm and cnn. IEEE Access, 8, 166026-166037.

[5] Xu, D., Chen, Y., Zhao, H., Li, Y., & Yuan, Y. (2021). TranAD: Deep Transformer Networks for Anomaly Detection. arXiv preprint arXiv:2110.04174.