

Received 26 November 2022, accepted 11 December 2022, date of publication 19 December 2022, date of current version 22 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3230352

RESEARCH ARTICLE

Machine Learning-Based Anomaly Detection for Multivariate Time Series With Correlation Dependency

SHASHANK CHAUHAN¹ AND SUDONG LEE¹

School of Industrial Engineering, University of Ulsan, Ulsan 44610, South Korea

Corresponding author: Sudong Lee (sudonglee@ulsan.ac.kr)

This work was supported by a grant of the National Research Foundation of Korea (Project Number: 2020R1F1A1072381).

ABSTRACT Recent advances in data collection facilitate the acquisition of large quantities of multivariate time series (MTS) data from various real-world systems. Anomaly detection in high-dimensional MTS data is essential to improving the productivity and safety of such systems; however, capturing the complex intercorrelations between different pairs of time series related to anomalous patterns is challenging. In this study, two different anomaly detection problems—mean shift and structural change—were defined based on the correlation dependency of MTS. Existing algorithms were experimentally analyzed and compared based on their correlation dependency encoding methods using synthetic datasets, with the results revealing that the explicit encoding of correlation dependency improves the predictive performance of anomaly detection in MTS data.

INDEX TERMS Anomaly detection, multivariate time series, correlation dependency.

I. INTRODUCTION

Industrial digitalization and advances in data collection technologies facilitate the acquisition of MTS data from various real-world systems. Anomaly detection in MTS refers to the identification of abnormal patterns in the collected data at particular time steps. The timely detection of anomalies improves the productivity and safety of systems by preventing possible breakdowns or accidents. Machine learning (ML) has been used to solve this problem in various areas, including manufacturing [1], [2], [3], [4], fraud detection [5], [6], [7], medical treatment [8], [9], [10], networks [11], [12], [13], and security [14], [15], [16].

Constructing an ML model for anomaly detection in MTS data is challenging because it requires capturing two dependencies in data. First, time series data exhibit “temporal dependency,” in which time series observations are interdependent according to temporal order because data points are affected by the data observed at previous time steps. Traditional anomaly detection methods such as the

histogram-based outlier score (HBOS) [17], isolation forest (IF) [18], local outlier factor (LOF) [19], and autoencoders (AEs) [20] assume that observations at different time steps are independent and identically distributed. Therefore, these methods cannot effectively encode temporal dependency. Recently, long short-term memory (LSTM) [21], [22], gated recurrent unit (GRU) [23], [24], and attention-based models [25], [26] have been used to address this problem. Second, the variables in MTS data are intercorrelated under what is referred to as “correlation dependency.” Traditional range-based anomaly detection approaches independently monitor each time series to determine whether the observed values lie between the lower and upper bounds and aggregate the results for a decision. However, complex systems such as manufacturing plants, power grids, and communication networks comprise multiple intercorrelated components, making it necessary to exploit cross-variable dependency to model the system’s normal conditions [27].

Two major problems with respect to correlation dependency have been defined for anomaly detection in MTS data. The first problem type involves the “mean shift” that occurs in single or multiple variables because of an unexpected

The associate editor coordinating the review of this manuscript and approving it for publication was Li He¹.

change in the data generation process. A significant mean shift in a single variable can be monitored in a univariate manner. However, if the variables are intercorrelated variation in one variable transfers to the others along the correlation structure. Therefore, anomalous patterns can be detected in advance if the model exhibits correlation dependency. The second problem is “structural change,” which occurs in the relationship between variables. In MTS data, events within individual variables and variations in relationships between variables affect the distribution of the multivariate observations. Explicit monitoring of the variation in a correlation structure is essential for anomaly detection in a complex system that can exhibit accumulated relationship changes.

Despite its significance, correlation dependency received less attention than temporal dependency. We have identified two primary reasons for this. First, before recent advances in the Industrial Internet of Things (IIoT), there had been a lack of data containing information on the interconnections between sensors, actuators, and other system instruments. Although machine learning methods have been developed to overcome the problem of data scarcity, large quantities of multivariate data with complex relationships are now being collected from various industrial sites. Second, machine learning methodologies, are generally worse at reflecting correlation dependency than they are at reflecting temporal dependency. Recently, we found that Graph Neural Networks (GNNs) [28], [29], [30] can address correlation dependency in MTS through the explicit mapping of structures using graphs. The multi-scale convolutional recurrent encoder-decoder (MSCRED) proposed by Zhang et al. [30] and the Graph Deviation Network (GDN) proposed by Deng and Hooi [28] are representative explicit methods that utilize graph structures to explicitly represent correlation dependencies in MTS data.

In this paper, we provide an approach to understanding MTS anomaly detection problems using correlation dependency. We first define two problem types related to correlation dependency and then comprehensively analyze the characteristics and effectiveness of the explicit incorporation of correlation dependency into an anomaly detection model. The contributions of this study are summarized as follows:

- We define the two problem types in MTS anomaly detection related to correlation dependency, i.e., mean shift and structural change, using rigorous definitions and examples.
- We propose a categorization of machine learning-based MTS anomaly detection models according to the correlation dependency encoding method.
- We comprehensively compare the performances and characteristics of the prevalent algorithms for mean shift and structural change at various correlation levels.

The remainder of this paper is organized as follows: Section 2 briefly introduces related works. Section 3 proposes a categorization of MTS anomaly detection algorithms based on correlation dependency by discussing ten relevant

algorithms. Section 4 describes our experimental study and results. Section 5 presents our conclusions.

II. RELATED WORKS

Here, we summarize the works related to our research. First, we review the previous work on MTS anomaly detection with correlation dependency. We then introduce the use of the linear Gaussian model as a graphical model for representing multivariate data correlation structures. Finally, we define the two MTS anomaly detection problem types related to correlation dependency, i.e., mean shift and structural change.

A. ANOMALY DETECTION IN MTS WITH CORRELATION DEPENDENCY

The past few decades have seen a growing corpus of research on machine learning methods for anomaly detection in MTS data. As noted in the introduction, correlation dependency has received relatively little attention owing to a lack of data and methodologies. Recently, the big data collected by the IIoT and the rise of advanced machine learning techniques have allowed more attention to be focused on the problem. In particular, graph-based deep learning methods have proven to be capable of effectively encoding the correlation dependencies of MTS data in models. Yu et al. [31] used Graph Convolution Networks (GCNs) to learn spatial-temporal correlations in time series forecasting problems in the traffic domain. Zhao et al. [29] proposed a prediction-based GNN combining a feature-oriented Graph Attention (GAT) layer and a time-oriented GAT to capture both spatial and temporal dependencies in MTS. Zhang et al. [30] proposed the Multi-Scale Convolutional Recurrent Encoder-Decoder (MSCRED) for anomaly detection in MTS data. MSCRED encodes correlation dependency using a convolutional encoder and decoder with a signature matrix and incorporates temporal patterns using attention-based Convolutional Long-Short Term Memory (ConvLSTM) networks. It constructs the signature matrices to characterize multiple levels of system status across time steps, which are used to indicate the severity of different abnormal incidents. Dang et al. proposed the Graph Deviation Network (GDN) [28], a novel attention-based GNN approach that learns graphs of the correlation dependencies between variables, and identifies and explains deviations from these relationships. Both algorithms have demonstrated excellent performance in real-world applications of MTS anomaly detection using graph-based explicit correlation dependency encoding.

B. LINEAR GAUSSIAN MODEL

Multivariate time series in real-world systems often exhibit complex correlation structures among multiple temporal sequences [32]. Various metrics can be used to analyze the correlation between any two time series by measuring the degree to which one series evolves relative to another. The Pearson correlation coefficient is a typical correlation measure of the linear dependency between any two time series $X = [x_1, x_2, \dots, x_T]$ and $Y = [y_1, y_2, \dots, y_T]$ is

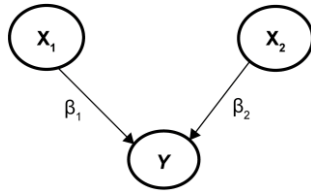


FIGURE 1. Illustration of an LGM.

defined as follows:

$$\rho_{X,Y} = \frac{cov(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^T (x_t - \bar{x})^2} \sqrt{\sum_{t=1}^T (y_t - \bar{y})^2}}, \quad (1)$$

where \bar{x} and \bar{y} are the mean values of X and Y , respectively. However, there are limits to the effectiveness of representing the overall correlation structure of MTS data using this type of correlation measure alone.

The correlation structure of MTS can also be modeled using a graphical model that provides a graph-based representation of the relations between the variables. A Bayesian network (BN) is a probabilistic graphical model that represents a set of conditional probability distributions of different variables using a directed acyclic graph. In particular, such graphical models represent the correlation structure of multiple variables and describe their direct influence on each other in specific directions [33]. A linear Gaussian model (LGM) is a special type of BN that can model relationships between multiple continuous variables using conditional linear gaussian distributions. One significant advantage of LGMs is that they can model the causal relationships between variables once the graph structure is determined, enabling the modeling of real-world processes that generate MTS data [34], [35].

Figure 1 illustrates an LGM for MTS data. Each node in the graph denotes a random variable; each directed edge connecting a parent and child node represents the causal effect between the two variables. In Figure 1, the time series for Y is generated by the conditional linear Gaussian distributions of X_1 and X_2 as follows:

$$P(Y | X_1, X_2) = N(\beta_0 + \beta_1 X_1 + \beta_2 X_2, \sigma^2), \quad (2)$$

where σ denotes the random noise and the linear coefficient β represents the effect of the parent node on the child node. The variables in the root nodes without any parents such as X_1 and X_2 are independently generated by the corresponding univariate Gaussian distribution $N(\mu, \sigma^2)$. Thus, a directed graph generated using LGM can represent a correlation structure of multiple variables via conditional distributions.

C. TYPES OF ANOMALIES IN MTS WITH CORRELATION

In this paper, we define two types of MTS anomaly detection problems, mean shift and structural change, in terms of correlation dependency. LGMs are used to demonstrate the defined problem types and to generate synthetic datasets.

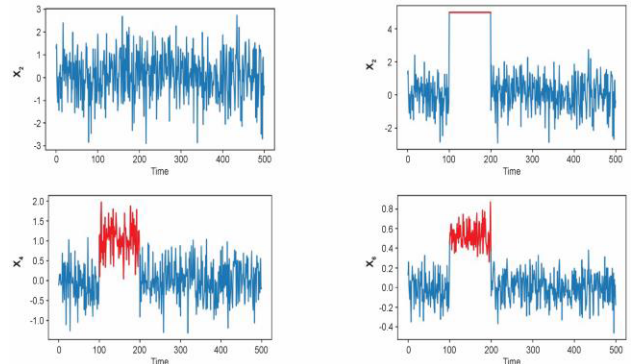
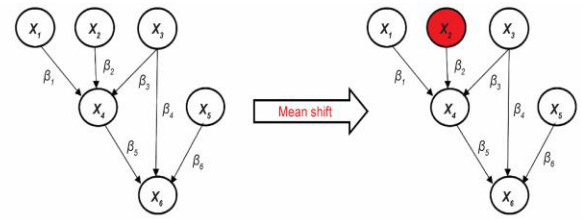


FIGURE 2. Illustration of mean shift anomaly using an LGM.

1) MEAN SHIFT ANOMALY

Time series data in a steady state are assumed to be generated from a normal distribution with a constant mean and variance. Abnormalities owing to mechanical failure, sensor malfunction, human error, or intrusion attacks can significantly shift the mean value of the distribution. We define this problem as a “mean shift” as follows:

Definition 1: Mean shift anomaly.

A mean shift anomaly is an anomalous pattern in the mean value of single or multiple variables caused by an unexpected change in the data generation process.

Because a mean shift that occurs in one or multiple variables propagates to the correlated variables, understanding the correlation structure can help detect such anomalous patterns more quickly and accurately. Figure 2 demonstrates the effect of one variable’s mean shift on other variables according to the correlation structure. Assume that a significant mean shift has occurred in X_2 . This change propagates directly to X_4 , which is a child node of X_2 , and affects X_6 through the relationship between X_4 and X_6 .

2) STRUCTURAL CHANGE ANOMALY

A multivariate joint probability distribution is determined by the distribution of each variable and the conditional probability between different variables. Whereas the mean shift relates to the variation of individual variables, an anomaly pattern occurring in a multivariate system can also reflect changes in the conditional distributions. We define this problem as “structural change” as follows:

Definition 2: Structural change anomaly.

A structural change anomaly is an anomalous pattern caused by a significant change in the relationship between variables rather than in the variables themselves.

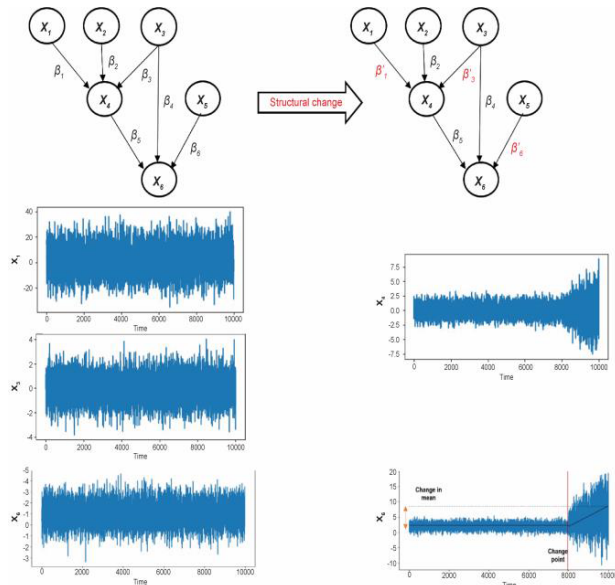


FIGURE 3. Illustration of structural change anomaly using an LGM.

For instance, a drilling machine can produce or enlarge holes in a solid material using drills. Input parameters such as cutting speed, feed, and drill diameter determine the process output. However, as the drill bit gradually wears out, output values such as the diameters or depths of holes vary despite identical input parameter settings. If the magnitude of a structural change increases or its effect accumulates over time, anomalies can arise in the multivariate system. Figure 3 illustrates these structural changes using an LGM. As the coefficients β_1 , β_3 , and β_6 transform to β'_1 , β'_3 , and β'_6 , respectively, the distributions of X_4 and X_6 vary as a result of the structural changes.

III. METHODS

Here, the existing anomaly detection methods are categorized to compare their characteristics and performance for MTS data with correlation dependency. We first divide the methods into univariate and multivariate methods according to whether or not the method inputs multiple variables individually. Because a univariate method independently processes individual time series, the correlation dependency cannot be reflected in the model. By contrast, a multivariate method simultaneously inputs multiple time series. Multivariate methods are further categorized into implicit and explicit methods in terms of how they encode correlation dependency. An implicit method assumes that the input variables are independent, i.e., without any explicit encoding of the relationship between variables. By contrast, an explicit method explicitly models the correlation structure using a representation such as a correlation matrix or graphical model. The characteristics of the methods we assessed are summarized in Table 1.

A. UNIVARIATE METHODS

1) HISTOGRAM-BASED OUTLIER SCORE (HBOS)

HBOS [17] measures the multivariate abnormality of a data point in terms of individual variables by summing the

univariate anomaly scores. It constructs a histogram for each variable in which the inverse height indicates the outlier score of a variable's data point. Because histograms are easy to construct, HBOS is a computationally efficient unsupervised anomaly detection method.

2) INTERQUARTILE RANGE (IQR)

IQR [36] is a range-based anomaly detection method in which a range is defined using upper and lower bounds and three quartiles to detect out-of-range anomalies in a single time series.

B. MULTIVARIATE METHODS

1) IMPLICIT METHODS

a: ISOLATION FOREST (IF)

IF [18] randomly constructs multiple classification trees that isolate each training observation from the others. In recursive partitioning using trees, the path length from the root to the leaf node measures the observation's abnormality because a forest of random trees collectively produces shorter path lengths for anomalies.

b: LOCAL OUTLIER FACTOR (LOF)

LOF [19] is a distance-based anomaly detection method that uses local densities in multivariate data to detect anomalies. The k -nearest neighbors for each data point are evaluated to calculate the local densities of all data points, or local reachability densities (LRDs). The anomaly score for each data point is then estimated by comparing its LRD values with those of its successive k -nearest neighbors.

c: AUTOENCODER (AE)

AE [20] is an artificial neural network that learns a compressed representation of input data using reconstruction. As it is assumed that an AE model trained on a normal dataset will fail to reconstruct unseen anomalies, the anomaly score is estimated using the reconstruction error; this approach is known as reconstruction-based anomaly detection.

d: LSTM AUTOENCODER (LSTM-AE)

LSTM-AE [37] is one of the most successful methods for MTS anomaly detection. The LSTM structure in an AE represents the temporal dependency of time-series data.

e: DEEP AUTOENCODING GAUSSIAN MIXTURE MODEL (DAGMM)

DAGMM [38] utilizes a deep autoencoder to generate a low-dimensional representation and reconstruction error for each input data point, which is further fed into a Gaussian Mixture Model (GMM). Instead of using decoupled two-stage training and the standard Expectation-Maximization algorithm, DAGMM jointly optimizes the parameters of the deep autoencoder and the mixture model simultaneously in an end-to-end fashion, leveraging a separate estimation network to facilitate the parameter learning of the mixture model.

TABLE 1. Summary of various anomaly detection methods.

Category		Method	Detection Strategy	Intercorrelation Encoding						
Univariate		HBOS	Histogram	None						
		IQR	Range	None						
Multivariate		Implicit			IF	Tree	None			
					LOF	Density	None			
					AE	Reconstruction	None			
					LSTM-AE	Reconstruction	None			
					DAGMM	Reconstruction	None			
					enGAN	Discrimination	None			
					Explicit			MSCRED	Reconstruction	Signature matrix
								GDN	Forecasting	Graph

f: GAN ENSEMBLE (enGAN)

enGAN [39] models an ensemble of GANs that utilize multiple generators and discriminators that are randomly paired and trained via adversarial training. Each discriminator obtains feedback from multiple discriminators, each of which is fed training samples from multiple generators. The anomaly score is computed as the average of the anomaly scores from all the generator-discriminator pairs. The authors found that enGAN can better model the distribution of normal data, which allows it to outperform a single GAN in detecting anomalies.

2) EXPLICIT METHODS

a: MULTI-SCALE CONVOLUTIONAL RECURRENT ENCODER-DECODER (MSCRED)

MSCRED [30] constructs multi-scale signature matrices to characterize multiple levels of system status at different time steps. A convolutional encoder incorporates the correlations between variables using the provided signature matrices. Convolutional LSTM (ConvLSTM) is employed to capture temporal dependency. Based on the feature maps, a convolutional decoder reconstructs the input signature matrices, and the residual measures the anomaly score.

b: GRAPH DEVIATION NETWORK (GDN)

GDN [28] represents cross-variable correlation dependency using a graphical model. A GDN model learns the graph structure of variables and predicts a variable’s behavior using an attention function over its neighbors in the graph. Furthermore, graph deviation scoring identifies anomalies that deviate from the learned relationships in the graph.

IV. EXPERIMENTS

This section describes our experimental assessments and discusses the results.

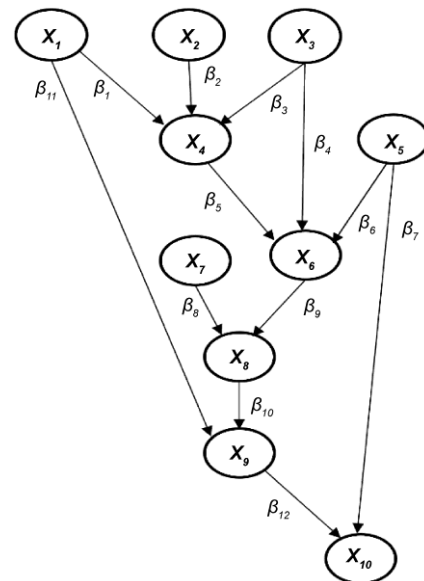


FIGURE 4. LGM model used for synthetic data generation in experiments.

A. DATASETS

The synthetic datasets for mean shift and structural change were generated using an LGM, as shown in Figure 4. The LGM comprised 10 variables, namely, X_1 to X_{10} . The root node variables X_1 , X_2 , X_3 and X_5 were independently generated from $N(0, 1)$. Given the parent variable values, the child node variables were generated from the corresponding conditional linear Gaussian distribution with the linear regression coefficients β and unit variance.

1) MEAN SHIFT

A mean shift, which significantly shifts the mean value of X by Δ from the starting time point s to the end time point e ,

TABLE 2. Experimental results of mean shift detection in the low correlation dataset.

Mean Shift	Low					Moderate					High				
	Methods	Accuracy	Precision	Recall	F1-score	AUC	Accuracy	Precision	Recall	F1-score	AUC	Accuracy	Precision	Recall	F1-score
HBOS	0.711	0.123	0.071	0.091	-	0.754	0.156	0.122	0.137	-	0.764	0.266	0.356	0.304	-
IQR	0.698	0.174	0.226	0.197	-	0.706	0.223	0.321	0.263	-	0.726	0.445	0.455	0.452	-
IF	0.806	0.422	0.422	0.422	0.212	0.824	0.556	0.559	0.557	0.256	0.854	0.603	0.699	0.647	0.315
LOF	0.807	0.411	0.411	0.411	0.223	0.803	0.456	0.466	0.461	0.269	0.857	0.501	0.501	0.501	0.306
AE	0.639	0.706	0.366	0.536	0.559	0.654	0.711	0.425	0.596	0.603	0.694	0.765	0.556	0.715	0.625
LSTM-AE	0.826	0.974	0.486	0.648	0.643	0.839	0.974	0.523	0.684	0.669	0.843	0.988	0.602	0.748	0.689
DAGMM	0.856	0.699	0.426	0.529	0.633	0.874	0.682	0.468	0.555	0.647	0.899	0.732	0.499	0.614	0.736
enGAN	0.788	0.614	0.596	0.605	0.658	0.866	0.648	0.601	0.624	0.693	0.887	0.758	0.624	0.685	0.701
MSCRED	0.854	0.699	0.779	0.743	0.673	0.907	0.701	0.801	0.756	0.701	0.893	0.795	0.805	0.791	0.756
GDN	0.901	0.705	0.856	0.773	0.745	0.928	0.725	0.889	0.799	0.806	0.924	0.755	0.905	0.823	0.822

is formulated as

$$X(t) \sim \begin{cases} N(\mu, \sigma^2), & \text{for } 1 \leq t \leq s \\ N(\mu + \Delta, \sigma^2), & \text{for } s \leq t \leq e \\ N(\mu, \sigma^2), & \text{for } e < t \leq T, \end{cases} \quad (3)$$

where $t = 1, 2, \dots, T$. In our experiments, we generated 10,000 observations (*i.e.*, $T = 10,000$). We assumed that a mean shift occurred at $s = 6,000$, and the mean value of X_1 changed from the initial value $\mu = 1$ to $\mu + \Delta$ until $e = 8,000$. Thereafter, the mean value returned to the initial value, $\mu = 1$. The results were compared on various mean shift levels, with $\Delta = 1, 5$, and 10 . The mean shift in X_1 propagated to the other variables along the edges of the LGM, with each edge representing the correlation between the parent and child nodes via the linear regression coefficient β . The effect of mean shift on the occurrence of anomalies in the entire multivariate system depended on the magnitude of the correlation between variables. For comparison, three different correlation levels—low, moderate, and high—were set as follows:

- Low correlation: $\beta = 0.1$
- Moderate correlation: $\beta = 2$
- High correlation: $\beta = 5$

2) STRUCTURAL CHANGE

A structural change occurs in the causal relationship of one or more pairs of variables. An LGM represents the structural change between two variables by applying different coefficient values β' for the initial coefficient β after an event. A structural change increases the value of β by Δ per unit time after the event time s as follows:

$$\beta'(t) = \begin{cases} \beta & 0 < t < s \\ \beta + \Delta \times t & s \leq t \leq e \\ \beta & e < t \leq T, \end{cases} \quad (4)$$

where $t = 1, 2, \dots, T$. In the initial state until $s = 6,000$, data were generated by the LGM in Figure 4 with β_1, β_3 , and β_6 set to 2, 3, and 4, respectively, and the other coefficients

set to 1. It was assumed that structural changes occurred in β_1, β_3 , and β_6 from $s = 6,000$ to $e = 8,000$ as the three coefficient values increased by $\Delta = 0.01$ at each time step.

B. EVALUATION

Here, we evaluate the anomaly detection methods using synthetic datasets generated for mean shift and structural change. The datasets were standardized to zero mean and unit variance and then divided into training and test data in a 60:40 ratio (*i.e.*, $T = 1-6,000$ for training and $T = 6,001-10,000$ for testing, respectively). Out of the training data, 20% ($T = 4,801-6,000$) were used as validation data to determine the anomaly score threshold. We note that all of the anomaly detection methods used in the experiments require normal data alone for training because it was unfeasible to obtain a sufficient quantity of labeled abnormal data from the industrial sites. The standard evaluation metrics—accuracy, precision, recall, F1 score, and area under the receiver operating curve (AUC)—were adopted for the evaluation of anomaly detection performance. All of the methods were implemented using Python 3.7. HBOS, IQR, IF, and LOF were trained using Scikit-learn [40], and AE, LSTM-AE, DAGMM, enGAN, MSCRED, and GDN were implemented using the TensorFlow [41] and Keras [42] frameworks. All experiments were conducted using an Intel®Core™i5-10400 CPU @ 2.90GHz, 24GB RAM, and Windows 10 (64-bit).

C. EXPERIMENTAL RESULTS

1) MEAN SHIFT DETECTION

Tables 2–4 present the anomaly detection performances of the respective methods at various levels of correlation and mean shift. The best performance for each case is highlighted in bold. Regardless of the level of correlation and mean shift, the explicit methods performed best and the univariate methods performed worst, as the latter assumes that all variables are independent and therefore cannot effectively encode correlation dependency. However, LSTM-AE performed well relative to the other implicit methods at all correlation levels and performed particularly well on highly correlated data.

TABLE 3. Experimental results of mean shift detection in the moderate correlation dataset.

Mean Shift	Low					Moderate					High				
	Methods	Accuracy	Precision	Recall	F1-score	AUC	Accuracy	Precision	Recall	F1-score	AUC	Accuracy	Precision	Recall	F1-score
HBOS	0.711	0.156	0.126	0.139	-	0.756	0.198	0.136	0.161	-	0.801	0.364	0.356	0.361	-
IQR	0.714	0.184	0.233	0.206	-	0.703	0.226	0.322	0.266	-	0.751	0.554	0.477	0.513	-
IF	0.816	0.456	0.456	0.456	0.241	0.868	0.589	0.559	0.574	0.266	0.866	0.665	0.714	0.689	0.411
LOF	0.823	0.411	0.411	0.411	0.255	0.804	0.477	0.471	0.474	0.278	0.853	0.556	0.601	0.578	0.356
AE	0.647	0.724	0.378	0.549	0.566	0.705	0.736	0.412	0.574	0.579	0.746	0.896	0.588	0.741	0.696
LSTM-AE	0.845	0.963	0.478	0.639	0.669	0.856	0.974	0.566	0.716	0.705	0.858	0.988	0.663	0.794	0.742
DAGMM	0.798	0.472	0.472	0.472	0.614	0.844	0.509	0.489	0.499	0.697	0.851	0.759	0.771	0.741	0.769
enGAN	0.804	0.801	0.658	0.722	0.701	0.811	0.811	0.678	0.738	0.714	0.888	0.824	0.707	0.761	0.724
MSCRED	0.906	0.854	0.688	0.759	0.711	0.921	0.848	0.726	0.76	0.736	0.893	0.855	0.765	0.778	0.778
GDN	0.901	0.815	0.866	0.826	0.854	0.921	0.825	0.889	0.845	0.877	0.921	0.855	0.893	0.861	0.905

TABLE 4. Experimental results of mean shift detection in the high correlation dataset.

Mean Shift	Low					Moderate					High				
	Methods	Accuracy	Precision	Recall	F1-score	AUC	Accuracy	Precision	Recall	F1-score	AUC	Accuracy	Precision	Recall	F1-score
HBOS	0.736	0.226	0.198	0.211	-	0.754	0.256	0.213	0.233	-	0.801	0.355	0.315	0.334	-
IQR	0.686	0.223	0.266	0.243	-	0.706	0.366	0.478	0.415	-	0.756	0.566	0.489	0.525	-
IF	0.832	0.469	0.479	0.474	0.241	0.861	0.601	0.611	0.606	0.299	0.882	0.699	0.745	0.721	0.436
LOF	0.824	0.411	0.466	0.437	0.255	0.809	0.588	0.556	0.572	0.296	0.871	0.554	0.611	0.581	0.378
AE	0.644	0.885	0.498	0.665	0.566	0.736	0.874	0.455	0.574	0.601	0.789	0.901	0.601	0.751	0.701
LSTM-AE	0.843	0.934	0.534	0.722	0.709	0.874	0.977	0.614	0.754	0.745	0.883	0.988	0.699	0.819	0.789
DAGMM	0.887	0.579	0.904	0.706	0.701	0.901	0.604	0.855	0.708	0.724	0.921	0.601	0.869	0.711	0.735
enGAN	0.856	0.801	0.647	0.716	0.699	0.861	0.855	0.741	0.794	0.754	0.881	0.866	0.734	0.795	0.801
MSCRED	0.859	0.746	0.678	0.701	0.701	0.923	0.756	0.739	0.746	0.738	0.925	0.869	0.785	0.804	0.796
GDN	0.934	0.778	0.869	0.821	0.904	0.944	0.756	0.905	0.824	0.914	0.942	0.869	0.946	0.906	0.926

enGAN was also competitive compared to the other implicit models. As the correlation level increased, the explicit methods performed better than the implicit methods, as shown in Figure 5. MSCRED encodes the correlation dependency for normal data in signature matrices, allowing it to better identify anomalous variables as these signature matrices change. GDN explicitly learns correlation dependency using a graph, allowing it to accurately detect anomalies in MTS data with correlation. The results indicate that explicit multivariate approaches outperform other methods in the mean shift detection of MTS data with correlation dependency.

2) STRUCTURAL CHANGE DETECTION

Table 5 presents the experimental results for structural change detection. GDN achieved the highest F1 score of 0.915 and an AUC of 0.926. The other explicit method, MSCRED, achieved the second-best F1 score of 0.856 and an AUC of 0.869. GDN and MSCRED succeeded in accurate structural change detection by capturing very small changes in the cross-variable relationship through the explicit encoding of correlation dependency. By contrast, the univariate methods were insensitive to structural changes in MTS data. Among the implicit multivariate methods, LSTM-AE, DAGMM, and

TABLE 5. Experimental results of structural change detection.

Methods	Accuracy	Precision	Recall	F1-score	AUC
HBOS	0.641	0.234	0.221	0.227	-
IQR	0.458	0.333	0.102	0.156	-
IF	0.665	0.322	0.298	0.31	0.405
LOF	0.661	0.366	0.255	0.301	0.366
AE	0.714	0.666	0.654	0.66	0.618
LSTM-AE	0.746	0.696	0.845	0.751	0.701
DAGMM	0.788	0.701	0.825	0.758	0.754
enGAN	0.888	0.798	0.764	0.781	0.777
MSCRED	0.848	0.857	0.866	0.856	0.869
GDN	0.927	0.904	0.926	0.915	0.926

enGAN performed best. Although it lacks the explicit encoding of multivariate correlation dependency, LSTM-AE performed well by reflecting the temporality of data. DAGMM uses GMM to achieve better parameter learning of deep autoencoder models, allowing it to optimize network parameters and provide a robust threshold between normal and structural change anomalies. AE performed poorly in structural change detection relative to mean shift detection.

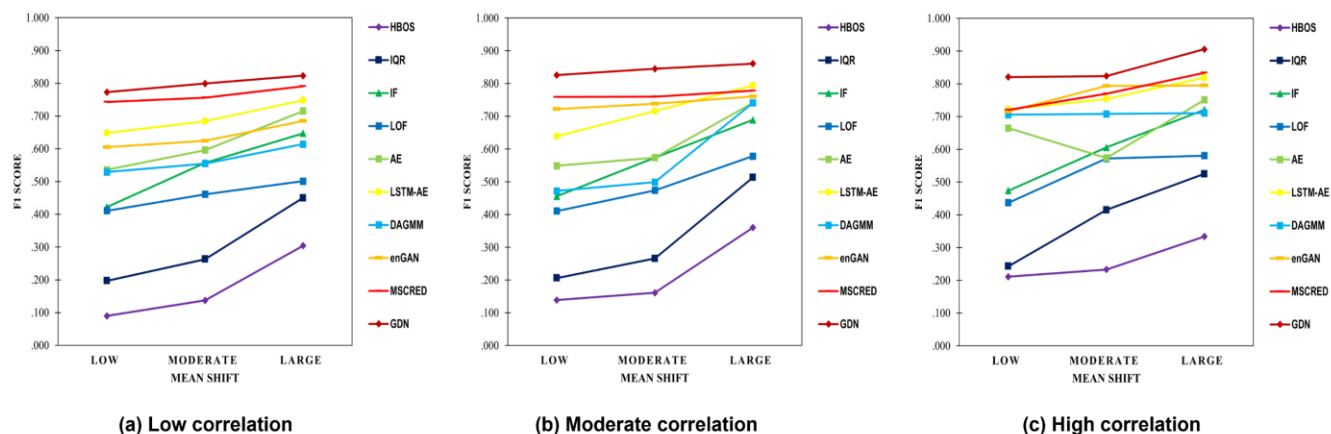


FIGURE 5. Comparison of F1 scores in the low, moderate, and high correlation datasets.

V. CONCLUSION

Understanding the correlation dependency between variables is essential for anomaly detection in MTS data. In this study, mean shift and structural change, which are MTS anomaly detection problems related to correlation dependency, were defined using rigorous definitions and examples. The existing anomaly detection methods were categorized based on the techniques used to incorporate correlation dependency, compared, and analyzed on a synthetic dataset generated using the LGM. The explicit models, GDN and MSCRED, outperformed the implicit and univariate methods in the detection of mean shifts and structural changes.

From the results, the following insights were obtained. First, the explicit incorporation of correlation dependency was found to improve mean-shift detection performance, a result that became more evident as the correlation between variables increased. Second, explicit models were able to effectively detect structural changes. Additionally, the encoding of temporal dependency improved the structural change detection performance, as indicated by the competitive results of LSTM-AE. Our experimental results also revealed the limitations of the implicit methods in capturing correlation dependency. We hope the reported results will motivate future research work to address this problem, which is highly relevant to real-world applications.

REFERENCES

- [1] P. Kamat and R. Sugandhi, "Anomaly detection for predictive maintenance in industry 4.0—A survey," in *Proc. E3S Web Conf.*, vol. 170, 2020, pp. 1–8, doi: [10.1051/E3SCONF/202017002007](https://doi.org/10.1051/E3SCONF/202017002007).
- [2] K. Kammerer, B. Hoppenstedt, R. Pryss, S. Stöckler, J. Allgaier, and M. Reichert, "Anomaly detections for manufacturing systems based on sensor data—Insights into two challenging real-world production settings," *Sensors*, vol. 19, no. 24, p. 5370, Dec. 2019, doi: [10.3390/S19245370](https://doi.org/10.3390/S19245370).
- [3] H. Kim and T. Shon, "Industrial network-based behavioral anomaly detection in AI-enabled smart manufacturing," *J. Supercomput.*, vol. 78, no. 11, pp. 13554–13563, Jul. 2022, doi: [10.1007/S11227-022-04408-4](https://doi.org/10.1007/S11227-022-04408-4).
- [4] B. Lindemann, F. Fesenmayr, N. Jazdi, and M. Weyrich, "Anomaly detection in discrete manufacturing using self-learning approaches," *Proc. CIRP*, vol. 79, pp. 313–318, Jan. 2019, doi: [10.1016/J.PROCIR.2019.02.073](https://doi.org/10.1016/J.PROCIR.2019.02.073).
- [5] J. P. A. Andrade, "A machine learning-based system for financial fraud detection," in *Proc. Encontro Nacional Inteligência Artificial Computacional*, 2021, pp. 165–176, doi: [10.5753/ENIAC.2021.18250](https://doi.org/10.5753/ENIAC.2021.18250).
- [6] E. Illeberi, Y. Sun, and Z. Wang, "A machine learning based credit card fraud detection using the GA algorithm for feature selection," *J. Big Data*, vol. 9, no. 1, pp. 1–17, Dec. 2022, doi: [10.1186/S40537-022-00573-8](https://doi.org/10.1186/S40537-022-00573-8).
- [7] M. Lokanan, V. Tran, and N. H. Vuong, "Detecting anomalies in financial statements using machine learning algorithm: The case of Vietnamese listed firms," *Asian J. Accounting Res.*, vol. 4, no. 2, pp. 181–201, Oct. 2019, doi: [10.1108/AJAR-09-2018-0032](https://doi.org/10.1108/AJAR-09-2018-0032).
- [8] T. Fernando, H. Gammulle, S. Denman, S. Sridharan, and C. Fookes, "Deep learning for medical anomaly detection—A survey," *ACM Comput. Surv.*, vol. 54, no. 7, pp. 1–37, Sep. 2022, doi: [10.1145/3464423](https://doi.org/10.1145/3464423).
- [9] G. D. Clifford, C. Liu, B. Moody, D. Springer, I. Silva, Q. Li, and R. G. Mark, "Classification of normal/abnormal heart sound recordings: The physionet/computing in cardiology challenge 2016," in *Proc. Comput. Cardiol. Conf. (CinC)*, vol. 43, Sep. 2016, pp. 609–612, doi: [10.22489/CINC.2016.179-154](https://doi.org/10.22489/CINC.2016.179-154).
- [10] C. Han, L. Rundo, K. Murao, T. Noguchi, Y. Shimahara, Z. Á. Milacski, S. Koshino, E. Sala, H. Nakayama, and S. Satoh, "MADGAN: Unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction," *BMC Bioinf.*, vol. 22, no. S2, pp. 1–20, Apr. 2021, doi: [10.1186/S12859-020-03936-1](https://doi.org/10.1186/S12859-020-03936-1).
- [11] K. Ding, Q. Zhou, H. Tong, and H. Liu, "Few-shot network anomaly detection via cross-network meta-learning," in *Proc. Web Conf.*, Apr. 2021, pp. 2448–2456, doi: [10.1145/3442381.3449922](https://doi.org/10.1145/3442381.3449922).
- [12] Q. Ma, C. Sun, and B. Cui, "A novel model for anomaly detection in network traffic based on support vector machine and clustering," *Secur. Commun. Netw.*, vol. 2021, pp. 1–11, Nov. 2021, doi: [10.1155/2021/2170788](https://doi.org/10.1155/2021/2170788).
- [13] B. A. Tama, L. Nkenyereye, and S. Lim, "A stacking-based deep neural network approach for effective network anomaly detection," *Comput., Mater. Continua*, vol. 66, no. 2, pp. 2217–2227, 2021, doi: [10.32604/CMC.2020.012432](https://doi.org/10.32604/CMC.2020.012432).
- [14] A. Diro, N. Chilamkurti, V. D. Nguyen, and W. Heyne, "A comprehensive study of anomaly detection schemes in IoT networks using machine learning algorithms," *Sensors*, vol. 21, no. 24, pp. 1–13, 2021, doi: [10.3390/S21248320](https://doi.org/10.3390/S21248320).
- [15] G. Raman, N. Somu, and A. P. Mathur, "A multilayer perceptron model for anomaly detection in water treatment plants," *Int. J. Crit. Infrastruct. Protection*, vol. 31, Dec. 2020, Art. no. 100393, doi: [10.1016/J.IJCIIP.2020.100393](https://doi.org/10.1016/J.IJCIIP.2020.100393).
- [16] J. Vávra, M. Hromada, L. Lukáš, and J. Dworzecki, "Adaptive anomaly detection system based on machine learning algorithms in an industrial control environment," *Int. J. Crit. Infrastruct. Protection*, vol. 34, Sep. 2021, Art. no. 100446, doi: [10.1016/J.IJCIIP.2021.100446](https://doi.org/10.1016/J.IJCIIP.2021.100446).

- [17] M. Goldstein and A. Dengel. (2012). *Histogram-Based Outlier Score (HBOS): A Fast Unsupervised Anomaly Detection Algorithm*. KI-2012 Poster Demo Track. [Online]. Available: [http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Histogram-based+Outlier+Score+\(+HBOS+\):+A+fast+Unsupervised+Anomaly+Detection+Algorithm#0](http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Histogram-based+Outlier+Score+(+HBOS+):+A+fast+Unsupervised+Anomaly+Detection+Algorithm#0)
- [18] F. T. Liu, K. M. Ting, and Z.-H. Zhou. (2008). *Isolation Forest ICDM08*. [Online]. Available: <https://cs.nju.edu.cn/zhouch/zhouch.files/publication/icdm08b.pdf?q=isolation-forest>
- [19] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2000, pp. 93–104, doi: [10.1145/342009.335388](https://doi.org/10.1145/342009.335388).
- [20] M. Schreyer, T. Sattarow, D. Borth, A. Dengel, and B. Reimer. "Detection of anomalies in large scale accounting data using deep autoencoder networks," 2017, *arXiv:1709.05254*.
- [21] Z. Zhao, C. Xu, and B. Li, "A LSTM-based anomaly detection model for log analysis," *J. Signal Process. Syst.*, vol. 93, no. 7, pp. 745–751, Jul. 2021, doi: [10.1007/S11265-021-01644-4](https://doi.org/10.1007/S11265-021-01644-4).
- [22] V. Q. Nguyen, L. Van Ma, and J. Kim, "LSTM-based anomaly detection on big data for smart factory monitoring," *J. Digit. Contents Soc.*, vol. 19, no. 4, pp. 789–799, 2018, doi: [10.9728/DCS.2018.19.4.789](https://doi.org/10.9728/DCS.2018.19.4.789).
- [23] L. Shen, Z. Li, and J. T. Kwok. "Time-series anomaly detection using temporal hierarchical one-class network," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 1–11.
- [24] Y. Guo, "Multidimensional time series anomaly detection: A GRU-based Gaussian mixture variational autoencoder approach," in *Proc. Asian Conf. Mach. Learn.*, vol. 95, 2001, pp. 97–112.
- [25] V. K. Kukkala, S. V. Thiruloga, and S. Pasricha, "LATTE: LSTM self-attention based anomaly detection in embedded automotive platforms," *ACM Trans. Embedded Comput. Syst.*, vol. 50, no. 5, pp. 1–23, 2021.
- [26] T. Tayeh, S. Aburakhia, R. Myers, and A. Shami, "An attention-based ConvLSTM autoencoder with dynamic thresholding for unsupervised anomaly detection in multivariate time series," *Mach. Learn. Knowl. Extraction*, vol. 4, no. 2, pp. 350–370, Apr. 2022, doi: [10.3390/MAKE4020015](https://doi.org/10.3390/MAKE4020015).
- [27] Y. Zhang, Z. Y. Dong, W. Kong, and K. Meng, "A composite anomaly detection system for data-driven power plant condition monitoring," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4390–4402, Jul. 2020, doi: [10.1109/TII.2019.2945366](https://doi.org/10.1109/TII.2019.2945366).
- [28] A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 5, pp. 4027–4035, doi: [10.1609/AAAI.V35I5.16523](https://doi.org/10.1609/AAAI.V35I5.16523).
- [29] H. Zhao, Y. Wang, J. Duan, C. Huang, D. Cao, Y. Tong, B. Xu, J. Bai, J. Tong, and Q. Zhang, "Multivariate time-series anomaly detection via graph attention network," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2020, pp. 841–850, doi: [10.1109/ICDM50108.2020.00093](https://doi.org/10.1109/ICDM50108.2020.00093).
- [30] C. Zhang, D. Song, Y. Chen, X. Feng, C. Lumezanu, W. Cheng, J. Ni, B. Zong, H. Chen, and N. V. Chawla, "A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data," in *Proc. 33rd AAAI Conf. Artif. Intell.*, 2019, pp. 1409–1416, doi: [10.1609/AAAI.V33I01.33011409](https://doi.org/10.1609/AAAI.V33I01.33011409).
- [31] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3634–3640, doi: [10.24963/IJCAI.2018/505](https://doi.org/10.24963/IJCAI.2018/505).
- [32] C. Hu, Y. Hu, and S. Seo, "A deep structural model for analyzing correlated multivariate time series," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2019, pp. 69–74, doi: [10.1109/ICMLA.2019.00020](https://doi.org/10.1109/ICMLA.2019.00020).
- [33] S. Epskamp, L. J. Waldorp, R. Möttus, and D. Borsboom, "The Gaussian graphical model in cross-sectional and time-series data," *Multivariate Behav. Res.*, vol. 53, no. 4, pp. 453–480, Jul. 2018, doi: [10.1080/00273171.2018.1454823](https://doi.org/10.1080/00273171.2018.1454823).
- [34] A. Maurya and M. Cheung, "Contrastive structured anomaly detection for Gaussian graphical models," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2018, pp. 736–739, doi: [10.1109/ASONAM.2018.8508475](https://doi.org/10.1109/ASONAM.2018.8508475).
- [35] U. Gather, M. Imhoff, and R. Fried, "Graphical models for multivariate time series from intensive care monitoring," *Statist. Med.*, vol. 21, no. 18, pp. 2685–2701, Sep. 2002, doi: [10.1002/SIM.1209](https://doi.org/10.1002/SIM.1209).
- [36] J. W. Tukey, "Exploratory data analysis," *Biometrics*, vol. 33, p. 688, Jan. 1977. [Online]. Available: <http://www.jstor.org/stable/2529486>
- [37] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "LSTM-based encoder–decoder for multi-sensor anomaly detection," 2016, *arXiv:1607.00148*.
- [38] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding Gaussian mixture model for unsupervised anomaly detection," in *Proc. 6th Int. Conf. Learn. Represent.*, 2018, pp. 1–19.
- [39] X. Han, X. Chen, and L. P. Liu, "GAN ensemble for anomaly detection," in *Proc. AAAI*, May 2021, vol. 35, no. 5, pp. 4090–4097, doi: [10.1609/AAAI.V35I5.16530](https://doi.org/10.1609/AAAI.V35I5.16530).
- [40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and M. Blondel, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.
- [41] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*.
- [42] F. Chollet, "Keras," 2015. [Online]. Available: <https://keras.io> and https://keras.io/getting_started/faq/#how-should-i-cite-keras



SHASHANK CHAUHAN received the B.Tech. degree in electronics and instrumentation engineering from the Vidya College of Engineering, Uttar Pradesh, India, in 2013. He is currently pursuing the M.S. degree in industrial engineering with the University of Ulsan, Ulsan, South Korea. He was a Process Control Officer at UFLEX Ltd., Noida, India. His current research interests include deep learning, the industrial Internet of Things (IIOT), and anomaly detection.



SUDONG LEE received the B.S. degree in industrial engineering from Hanyang University, Seoul, South Korea, in 2012, and the Ph.D. degree in industrial and management engineering from POSTECH, Pohang, South Korea, in 2018. He is currently an Assistant Professor with the Department of Industrial Engineering, University of Ulsan, Ulsan, South Korea. His research interests include machine learning, industrial AI, anomaly detection, causality, and data mining.

• • •