

A Review of Data-Driven Prediction of Undesirable Events in Offshore Oil Wells

—Based on the Public 3W Benchmark and Time-Series Deep Learning

Yiguan Liu, Xinru Liu, Jingru Yao, Yufan Wang, Beining Guo, Yicheng Wang*

College of Science, North China University of Science and Technology, Tangshan, China

Email: *546314305@qq.com

How to cite this paper: Liu, Y.G., Liu, X.R., Yao, J.R., Wang, Y.F., Guo, B.N. and Wang, Y.C. (2026) A Review of Data-Driven Prediction of Undesirable Events in Offshore Oil Wells. *Journal of Computer and Communications*, 14, 17-31.
<https://doi.org/10.4236/jcc.2026.145002>

Received: April 22, 2026

Accepted: May 22, 2026

Published: May 25, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The production and operation of offshore oil wells present typical characteristics of strong coupling, high nonlinearity, obvious time-varying behavior, and high operational risks. The occurrence of adverse events is rarely triggered by the over-limit behavior of a single variable; rather, it usually stems from the gradual deviation of multivariate correlation patterns in downhole systems from normal operating conditions. Traditional anomaly recognition methods based on threshold rules and manual experience are widely used in engineering practice and have practical value. Nevertheless, under complex operating-condition switching, weak precursor anomalies, and cross-well operational discrepancies, these methods are often affected by high false-alarm rates, poor transferability, and delayed responses. With the rapid development of technology, data-driven methods have gradually evolved into a core technical solution for anomaly detection and adverse-event prediction in offshore oil wells. Effective anomaly identification and early prediction based on production monitoring data have become important research directions in intelligent oilfield construction. Targeting research progress in predictive analytics for the oil and gas industry and state-of-the-art multivariate time-series anomaly detection methods, this paper conducts a systematic review of adverse-event prediction in offshore oil wells. It summarizes existing studies based on the public 3W benchmark dataset and analyzes key data challenges, including structural missingness, cross-well heterogeneity, class imbalance, and weak precursor features. Furthermore, this paper presents a review-oriented conceptual design of a missing-aware multi-scale time-frequency graph. Transformer framework tailored to offshore oil well scenarios, which is intended to support anomaly detection, adverse-event recognition, and early prediction in future empirical validation. From a design-rationale perspective, this study compares the proposed framework with conventional methods and existing

deep models, and discusses its potential advantages in four aspects: missing-aware input encoding, dual-domain time-frequency representation, physics-prior-initialized learnable graph structure, and multi-task learning.

Keywords

Offshore Oil Wells, Adverse Event Prediction, 3W Dataset, Deep Learning, Transformer, Data-Driven

1. Introduction

Offshore oil well production systems operate for long periods under the combined effects of high temperature, high pressure, long-distance transportation, and complex phase behavior. Significant coupling exists among downhole safety valves, choke valves, gas-lift systems, subsea manifolds, and surface processing facilities. When pressure fluctuation, flow instability, choke abnormality, severe blockage, or hydrate formation occurs in a local part of the system, it may not only reduce single-well productivity but also trigger severe production fluctuation, well shut-in, and even safety incidents. Because offshore operation and maintenance costs are extremely high, the production losses and maintenance costs caused by anomalies are often much higher than those of onshore wells. Therefore, the prediction of undesirable events in offshore oil wells has long been an important research topic.

From the perspective of industrial digitalization, oil and gas production is gradually shifting from an operation and maintenance mode that relies on experience and fixed rules to a proactive management mode based on monitoring data, predictive analytics, and intelligent early warning. Azmi [1], D’Almeida [2], and Tariq [3] present similar observations. With the advancement of digitalization, predictive analytics, condition monitoring, and intelligent warning based on multi-source monitoring data are becoming important directions for the development of oil well adverse-event detection and digital transformation.

The emergence of public data benchmarks has further promoted the development of data-driven research. The 3W Dataset released by Vargas provides a standardized benchmark for the study of rare undesirable events in offshore oil wells [4]. Subsequently, Marins [5], Machado [6], Fernandes [7], and Bayazitova [8] conducted studies on this benchmark from the perspectives of tree models, one-class classifiers, autoencoders, and deep recurrent networks, respectively, accelerating the transition of this field from experience-based diagnosis to data-driven analysis. Meanwhile, review studies in the oil and gas industry have also pointed out that predictive analytics, condition monitoring, and intelligent warning are becoming key supports for digital transformation [9].

Based on these observations and the special data structure of the dataset, this paper conducts a technology-oriented review of existing studies and proposes a design hypothesis to be validated, namely SDG-Former. Compared with existing

studies, its expected benefits come from missing-aware input, time-frequency dual-domain representation, physics-prior-initialized learnable graph structure, and a multi-task output mechanism [10].

2. Related Work

2.1. Task Definitions under the 3W Dataset Label Structure

Research on undesirable events in offshore oil wells is not a single anomaly detection problem, but involves three levels: anomaly detection, event recognition, and early prediction. This paper defines these tasks under the 3W label structure. The 3W data are organized as multivariate time-series instances. Version-1.0.0 of the 3W Dataset contains 1984 CSV instances and 8 process variables, and includes missing, frozen, and outlier characteristics from real industrial data.

An undesirable event refers to an operating condition that causes an oil well to deviate from normal production and may lead to production loss, increased maintenance cost, or safety risk. In the 3W label system, normal operation is coded as 0, steady-state undesirable events are usually coded as 1 - 8, transition states are coded as 101 - 108, and unknown states are coded as NaN. Among them, 101 - 108 can be mapped to the transient precursors of the corresponding steady-state events 1 - 8 (**Table 1**).

Anomaly detection corresponds to a binary classification task: given a time window, the model determines whether it deviates from the normal production pattern. A window containing only valid label 0 is labeled normal; a window containing steady-state event labels 1 - 8 or transition labels 101-108 is labeled anomalous. Windows containing a large proportion of NaN values are not used for supervised loss, or they participate only in unsupervised representation learning through masking.

Event recognition corresponds to a multi-class classification task: under the condition that an anomaly has been detected, the model further determines which type of undesirable event is more likely. Specifically, steady-state labels are used as classes, while transition label 100+e is mapped to the same event e so that the transient segment and the steady-state segment remain semantically consistent.

Table 1. Abnormal state types in the 3W dataset [4].

Code	Event Type	Number of Instances
0	Normal	597
1	Abrupt increase of BSW	129
2	Spurious closure of DHSV	38
3	Severe slugging	106
4	Flow instability	344
5	Rapid productivity loss	451
6	Rapid closure of PCK	221
7	Scaling in PCK	14
8	Hydrate in production line	84

Early prediction corresponds to an early-warning task: before the steady-state undesirable event begins, the model predicts event risk and type based on multivariate changes in the late normal stage or in the transition segment. Its core indicators are not ordinary window accuracy, but the lead time of the first valid alarm relative to the onset of the steady-state event, the early-detection rate, and the false-alarm rate.

2.2. Traditional Machine Learning Methods

From the perspective of traditional machine learning, the survey by Chandola *et al.* [11] and the review of time-series anomaly detection by Blázquez-García *et al.* [12] indicate that anomaly detection methods can be divided into four categories: statistical-boundary-based methods, density-estimation-based methods, distance-measurement-based methods, and normal-domain-learning methods. In offshore oil well scenarios, these methods have long occupied an important position. They require relatively low sample sizes and annotation quality, and they are also easier to transform into prediction rules in industrial systems.

The Local Outlier Factor proposed by Breunig *et al.* [13] identifies outliers based on local density differences. The Isolation Forest proposed by Liu *et al.* [14] amplifies the separability of abnormal points through random partitioning, whereas the one-class support vector machine proposed by Schölkopf *et al.* [15] realizes novelty detection by learning the support domain of normal samples. These models jointly form the baseline system for early 3W anomaly detection and also explain why Fernandes. [7] found that LOF and OCSVM remain competitive in multiple groups of experiments. However, the review of condition monitoring and prognostics by Jardine *et al.* [16] has pointed out that reliable diagnosis of complex industrial systems ultimately requires stronger time-series representation and higher-level anomaly detection.

2.3. Deep Learning Methods

Deep learning has promoted the transformation of industrial time-series anomaly detection from shallow discriminators to end-to-end representation learning. The review of time-series classification by Fawaz *et al.* [17] indicates that convolutional networks are better at extracting local patterns, recurrent networks are more suitable for modeling sequential dependence, and hybrid networks often show stronger adaptability to complex industrial data. Malhotra. [18] and Hundman. [19] demonstrate, respectively, from general industrial sequences and spacecraft telemetry scenarios, that LSTM-based prediction error or reconstruction error can effectively characterize abnormal behavior.

In unsupervised and semi-supervised directions, OmniAnomaly proposed by Su *et al.* [20], MAD-GAN proposed by Li *et al.* [21], DAGMM proposed by Zong *et al.* [22], and USAD proposed by Audibert *et al.* [23] have become representative methods for industrial multivariate anomaly detection. These methods improve the modeling capability of complex normal patterns from the perspectives of sto-

chastic recurrent modeling, generative adversarial learning, joint density estimation, and dual-autoencoder adversarial training, respectively. However, they also face problems such as training instability and sensitivity to missing values.

In recent years, Transformers and graph neural networks have further expanded the modeling boundary of anomaly detection. The Transformer proposed by Vaswani *et al.* [24] provides a general attention framework for long-dependency modeling. Informer proposed by Zhou *et al.* [25] reduces the computational overhead of long-sequence processing through sparse attention. Anomaly Transformer proposed by Xu *et al.* [26] uses association discrepancy to characterize anomalous points. TranAD proposed by Tuli *et al.* [27] enhances the efficiency and stability of multivariate anomaly diagnosis. DCdetector proposed by Yang *et al.* [28] strengthens anomaly representation from the perspective of contrastive learning and the modeling of interactions among variables. Compared with models that encode only along the time dimension, such methods are more suitable for industrial scenarios because sensors often correspond to clear physical coupling relationships.

For offshore oil wells, this means that future high-performance models are unlikely to arise from the extreme deepening of a single paradigm. Instead, they are more likely to come from systematic designs that integrate missing awareness, dual-domain representation, graph-temporal fusion, and multi-task prediction.

3. SDG-Former Conceptual Framework

3.1. Design Objectives and Overall Idea

This paper proposes SDG-Former (Missing-aware Multi-scale Time-frequency Graph Transformer) as a conceptual framework for predicting undesirable events in offshore oil wells. Its objective is to translate the data challenges of the 3W Dataset into verifiable model design hypotheses: the coexistence of long-term dependence and local abrupt changes, the coexistence of variable coupling and structural missingness, the coexistence of anomaly scarcity and event diversity, and the coexistence of engineering deployment requirements and deep-model complexity.

SDG-Former is constructed around the 3W Dataset and offshore oil well scenarios with four core modules: first, missing-aware input encoding based on sliding windows; second, dual-domain representation that preserves both time-domain evolution and frequency-domain disturbances; third, a graph-temporal Transformer backbone for learning variable associations and cross-scale dependencies; and fourth, multi-task output heads for anomaly detection, event recognition, and lead-time evaluation.

3.2. Missing-Aware Input

For the problem of inconsistent sensor availability in offshore oil wells, this paper does not simply impute all missing values as a single value. Instead, it constructs a value matrix X , a missing mask, and a time-interval input ΔX for each time

step. The purpose is to enable the model to distinguish between a genuinely low observation and the temporary unavailability of a variable, thereby avoiding the mislearning of missingness patterns as stable discriminative features.

Considering further that the available sensor sets may differ significantly across wells and files, this paper first generates structural missingness based on the available sensor set, and then uses a lightweight router to determine which parameter subspace the sample should enter. Compared with training a complete model separately for each missingness pattern, this shared-backbone and missingness-adaptive strategy preserves structural differences, avoids the unlimited expansion of the model pool, and is more suitable for the large sample size and significant heterogeneity of the 3W data.

3.3. Time-Frequency Dual-Domain Representation

If modeling is performed only in the time domain, short-term oscillations, phase shifts, and multi-period disturbances may easily be compressed into superficially smooth trends. If modeling is performed only in the frequency domain, the order and duration structure of anomaly development may be ignored. Therefore, this paper adopts a dual-domain representation strategy. On the time-domain side, the window sequence is divided into several temporal patches to reduce the computational overhead of long-sequence attention. On the frequency-domain side, short-time Fourier transform (STFT) and three-band statistics are applied to each variable window to explicitly retain frequency clues related to typical anomalies such as mechanical oscillation, choke fluctuation, and flow instability.

Temporal patches emphasize the process of anomaly development, whereas frequency-band statistics emphasize the form of abnormal vibration. These two representations are highly complementary in industrial scenarios. In offshore oil wells, for example, severe slugging and flow instability often show stronger periodicity in the frequency domain, whereas choke restriction and rapid productivity decline rely more heavily on time-domain trends and variable linkage relationships. Therefore, dual-domain representation is closer to the detection logic of practical engineering than single-domain modeling. Nevertheless, dual-domain representation remains a design hypothesis to be validated by experiments, rather than a conclusion empirically verified in this paper.

3.4. Graph-Temporal Transformer

At the backbone level, this paper adopts a three-stage structure consisting of graph structure learning, temporal self-attention, and cross-domain fusion. A learnable adjacency matrix and a graph initialized by physical priors are used to model the relationships among downhole pressure, upstream and downstream choke pressure, gas-lift flow rate, and related temperature variables. The graph structure is not directly fixed; instead, it is allowed to undergo lightweight adjustment during training so as to adapt to dynamic associations across different wells and operating conditions.

In the temporal dimension, a hybrid encoder combining sparse self-attention and local convolution is introduced. Sparse self-attention follows the efficient design of Informer [25] for long-sequence processing, while local convolution is used to preserve short-term abrupt changes and edge information. This avoids the over-smoothing of pure Transformer models in industrial short-window tasks and mitigates the insufficient long-distance dependency capture of pure convolutional models.

Finally, two gating units, namely sensor attention and frequency-band attention, are designed in the fusion layer. The former dynamically measures which variables are more important for the current judgment, while the latter distinguishes the frequency region in which the current anomaly is mainly expressed. Compared with Anomaly Transformer [26], which focuses on locating anomalies by association discrepancy, this design embeds interpretability into the model structure, thereby providing variable-level and frequency-band-level auxiliary evidence for subsequent engineering analysis.

3.5. Design Choices and Rationale

To demonstrate that the core modules of SDG-Former are not simply superimposed, this paper further summarizes the basis for key design choices and the hypotheses to be verified from three perspectives: data characteristics, model structure, and task objectives. Specifically, the window length mainly serves the needs of transition process coverage and real-time early warning; the time-frequency dual-domain representation is used to simultaneously characterize trend changes and oscillation disturbances; the learnable adjacency structure is used to express the potential correlations between multivariable sensors; and the multi-task loss is used to unify the objectives of anomaly detection, event recognition, and early prediction (Table 2). The relevant design choices, their basis, and the hypotheses to be verified are shown in Table 2.

Table 2. Key SDG-former design choices, rationale, and hypotheses to be validated.

Key Model Design	Design Rationale	Hypothesis to Be Validated
Window length of 120 - 192 sampling points	With a recording frequency of 1 Hz, this window corresponds to 2 - 3.2 minutes of operating time. This range balances transition-process coverage and computational cost [4].	Compared with overly short windows, it may better capture slow precursors; compared with overly long windows, it can reduce label mixing and real-time latency.
STFT + three-band statistics	Severe slugging and flow instability have oscillatory or periodic characteristics, and STFT is suitable for representing local frequency changes in non-stationary windows [17].	Time-domain trends and frequency-domain disturbances are complementary and may enhance the expression of weak precursors.
Learnable adjacency initialized by physical priors	Studies such as GDN show that learning sensor graph structures helps capture variable correlations and supports anomaly interpretation [25].	Physical priors may reduce the risk of arbitrary connections under small samples, while learnable edge weights can adapt to cross-well differences.

Continued

Multi-task loss	The 3W Dataset provides normal, transition, and steady-state event labels, and multi-task learning can use related tasks to share representations [4].	Jointly optimizing detection, recognition, and early prediction may improve task consistency, but loss weights must be adjusted to avoid objective conflicts.
-----------------	--	---

4. W Benchmark Validation Protocol and Implementation Recommendations

4.1. Introduction to the 3W Dataset

This paper uses the 3W Dataset [4] as the core benchmark. The dataset contains undesirable-event instances from naturally flowing offshore oil wells. It is multivariate, time-series-based, rare-anomaly-oriented, and clearly grounded in engineering scenarios. Therefore, it is suitable for anomaly detection, event classification, and early-warning research (Table 2). Existing studies based on the 3W Dataset indicate that its research value lies not only in the inclusion of multiple anomaly classes but also in its data characteristics: coexistence of missing and frozen values, significant class imbalance, and large differences in duration and dynamic characteristics among different instances (Table 3).

Table 3. Core features of the 3W dataset and their modeling implications.

Dimension	Main Content	Modeling Implication
Monitoring object	Multivariate monitoring sequences from naturally flowing offshore oil wells [4]	Suitable for studying anomaly detection and early warning in real production scenarios
Variable structure	Eight key variables, including pressure, temperature, and flow rate	Variable coupling and asynchronous responses need to be explicitly modeled
Label form	Normal, transition, and anomaly-stage information	Supports detection, classification, and lead-time evaluation
Data quality	Missing values, frozen values, noise, and source heterogeneity	Requires models with missingness robustness and generalization capability
Sample distribution	Rare anomalies and imbalanced categories	Requires reweighting and robust evaluation strategies during training

Data processing is performed by sliding-window segmentation. For missing values, neighboring-time imputation is used when the missing proportion is low; if missingness shows a structural distribution, the missing mask is preferentially retained. The sliding-window length is set to 120 - 192 sampling points according to anomaly duration, and the step size is set to half of the window length in order to balance the number of samples and temporal resolution.

The main experiment uses instances with all labels 0 - 8. Ablation experiments may separately report results using only real instances and using real, simulated, and hand-drawn instances together (Table 3). All sliding windows must inherit their source file ID and event ID. Training, validation, and test splits should use

file ID as the minimum grouping unit. If well IDs can be parsed from file names, well-level grouping should be preferred; otherwise, file-level GroupKFold should be adopted. Five-fold GroupKFold is recommended. All overlapping windows generated from the same file must appear in the same fold, and random splitting by window is prohibited.

4.2. Data Processing and Normalization Strategies

Missing-value processing follows the rule of “short-gap imputation, long-gap masking, and structural-missingness preservation.” When a continuous missing segment lasts no more than 10 seconds, linear interpolation or forward/backward neighboring imputation within the same file is used. When a continuous missing segment lasts more than 10 seconds but no more than 60 seconds, the imputed value is used only as a numerical placeholder, while the missing mask $M = 0$ and the time interval are retained. When a continuous missing segment lasts more than 60 seconds, or when the missing proportion of a variable in a file exceeds 30%, it is no longer treated as an ordinary gap but as a structurally unavailable variable. In this case, only zero-filled numerical placeholders, the missing mask, and the time interval are used, and the value is not regarded as a real observation.

Frozen values need to be treated separately because freezing is not ordinary missingness but information distortion caused by a sensor holding a constant value for a period of time. It is recommended to detect consecutive repeated values for each variable: if a variable remains exactly the same for no fewer than 60 consecutive sampling points, or if the absolute value of the first-order difference after normalization remains continuously lower than 10^{-4} , the interval is marked as a frozen segment. Frozen-segment values are not directly deleted. Instead, a frozen mask F is retained in the input, and the segment is down-weighted or excluded when calculating statistical features, STFT, and supervised loss. Normalization parameters must be fitted only on the training set within each cross-validation fold to avoid test leakage. For continuous variables such as pressure, temperature, and flow rate, robust standardization based on normal windows in the training set is recommended:

$$x' = \frac{x - \text{median}_{\text{train}}}{\text{IQR}_{\text{train}} + \delta}$$

where $\text{median}_{\text{train}}$ and $\text{IQR}_{\text{train}}$ are estimated only from normal windows in the training set, and epsilon is used to avoid division by zero. For variables with natural ranges, such as valve opening, min-max scaling based on the training-set range or physical range can be used. For binary or multi-valued valve states, the original discrete meaning is retained and a state mask is added. The validation and test sets only apply the parameters fitted on the training set and do not re-estimate them.

4.3. Baseline Model and Evaluation Metrics Setting Recommendations

Comparative models cover three categories. The first category consists of tradi-

tional machine learning baselines, including LOF [13], Isolation Forest [14], OCSVM [15], and Random Forest [5]. The second category consists of common deep learning baselines for industrial anomaly detection, including OmniAnomaly [20], MAD-GAN [21], DAGMM [22], USAD [23], and LSTM autoencoders [18] [19]. The third category consists of recent graph-temporal and Transformer baselines, including Anomaly Transformer [26], TranAD [27] DCdetector [28], and recent improved graph-temporal models [29].

The evaluation metrics include Accuracy, Recall, F1, ROC-AUC, and PR-AUC. For rare-anomaly scenarios, PR-AUC often better reflects the model's recognition quality for minority-class windows than ROC-AUC. For engineering early-warning tasks, the average early-detection ratio and the first detection time of the anomaly should also be added.

In terms of experimental settings and evaluation, the training stage adopts the AdamW optimizer, cosine-annealing learning rate, class reweighting, and early stopping. During inference, anomaly scores, variable attention, and frequency-band attention are retained for subsequent manual review and engineering interpretation. For edge-deployment scenarios, lower latency can be obtained by reducing the number of Transformer layers, shortening the patch length, and pruning the frequency-band branch.

5. Comparative Discussion Based on Existing Studies

Existing studies show that Random Forest methods have verified the feasibility of traditional supervised classification for clear fault patterns in early 3W tasks. Comparisons of one-class classifiers indicate that LOF and OCSVM remain competitive when normal samples are sufficient and anomalous samples are limited. Studies on deep recurrent networks and open-world learning [10] further show that deep models and open-set strategies can provide additional benefits under more complex anomaly-class conditions. The evaluation should not focus only on small improvements in a single metric; instead, it should emphasize stability under complex combinations, rare classes, and cross-well generalization. The following content compares only the method routes in existing literature with the design hypotheses of SDG-Former.

5.1. Discussion Compared with Traditional Methods

Traditional methods usually perform well in two types of scenarios: cases where anomaly boundaries are clear and statistical features are obvious, and cases where sample sizes are limited and deep models are prone to overfitting. Random Forest [5], LOF [13], and Isolation Forest [14] have clear advantages in this regard because they have low training costs, are friendly to small samples, and are less sensitive to hyperparameters and optimization procedures.

However, for adverse-event prediction in offshore oil wells, the limitations of traditional methods are also clear. They have difficulty directly using the continuously evolving temporal structure in transition stages and often rely on window-

level summary statistics. Their ability to express variable coupling is limited, and they may simplify multivariate linkage anomalies into several univariate shifts. When structural missingness and cross-well operating-condition changes exist, shallow models are more likely to mistake distribution habits in the training set for generalizable rules.

In comparison, the SDG-Former proposed in this paper has the following expected design advantages: it explicitly exposes weak precursors from both the time and frequency domains through dual-domain representation, thereby reducing reliance on manually designed statistics; it learns dynamic relationships among variables through graph-temporal fusion rather than treating each sensor as an independent channel; and it incorporates structural missingness into the model design through missingness routing instead of passively masking it during pre-processing.

5.2. Discussion Compared with Deep Learning Methods

Compared with deep learning models, this paper does not regard the reconstruction of normal patterns as the only objective. OmniAnomaly [20], MAD-GAN [21], DAGMM [22], and USAD [23] are representative models in industrial multivariate anomaly detection. However, such models usually rely on reconstruction error or energy functions to characterize anomalies. Therefore, when normal operating conditions themselves fluctuate substantially and normal patterns differ significantly across wells, threshold transfer becomes difficult.

Compared with graph-temporal and Transformer models, Anomaly Transformer [26], TranAD [27], DCdetector [28], and recent improved graph-temporal models have significantly improved multivariate anomaly detection performance. Nevertheless, most of them still assume fixed input dimensions and give insufficient consideration to changes in sensor availability. In addition, many models focus on detecting anomalous points, while unified modeling of event-type recognition and early-warning lead time remains insufficient.

5.3. Discussion Compared with Deep Learning Methods

For offshore oil well scenarios, the most interpretable result is often not a small improvement in a single AUC value, but whether the model can remain stable in cross-well testing, provide sufficiently early prediction before steady-state anomalies occur, and output key variables that are consistent with engineering knowledge (**Table 4**).

However, the proposed method also has limitations. Dual-domain representation and graph-temporal fusion increase implementation complexity and impose higher requirements on data preprocessing protocols. If the training data cover too few well types and operating conditions, the model may still be affected by distribution shift. Although multi-task learning improves task completeness, it may also introduce objective conflicts, requiring careful adjustment of loss weights and training strategies.

Table 4. Comparison of existing model routes and SDG-former design hypotheses.

Model Route	Representative Methods	Main Advantages	Main Limitations	Improvements in This Paper
Shallow anomaly detection	LOF [13], IF [14], OCSVM [15]	Fast training and friendly to small samples	Difficult to model long-term dependence and variable coupling	Introduces graph-temporal modeling and dual-domain representation
Supervised classification	RF, XGBoost, etc [5].	Strong capability for anomaly recognition	Strong label dependence and limited generalization	Enhances early prediction and robustness
Reconstruction-based deep models	OmniAnomaly [20], USAD [22], DAGMM [23]	Can learn complex normal patterns	Difficult threshold transfer and sensitivity to missingness	Uses multi-task discrimination instead of a single reconstruction path
Graph/Transformer models [30]	GDN [27], TranAD [28], DCdetector [30]	Can learn long dependencies and variable relationships	Many assume fixed input dimensions	Adds missingness routing and missing-aware modeling
Proposed method	SDG-Former	Balances detection, recognition, early warning, and interpretation	Relatively complex structure and high requirements for implementation protocols	Controls complexity through modular design

6. Conclusions and Prospects

This paper focuses on data-driven prediction of undesirable events in offshore oil wells. It reorganizes the research background, public data, related work, algorithm design, experimental settings, and discussion of results, and discusses the potential applicability of SDG-Former from the perspective of a conceptual framework. The difficulty of predicting undesirable events in offshore oil wells lies not only in the scarcity of anomalous samples, but also in complex variable coupling, obvious structural missingness, strong cross-well heterogeneity, and the need for early prediction of undesirable events.

Based on a systematic analysis of the 3W Dataset and related literature, the SDG-Former proposed in this paper forms an algorithmic design route that is more suitable for offshore oil well scenarios through missing-aware input, time-frequency dual-domain representation, a graph-temporal Transformer backbone, and a multi-task output mechanism. It provides a structured research framework that unifies anomaly detection and undesirable-event prediction.

Future research can proceed in three directions. First, self-supervised pretraining and foundation-modeling ideas can be introduced to improve transferability in small-sample and multi-well scenarios. Second, physical mechanisms, well-bore-flow constraints, and operating procedures can be incorporated into neural-network training to enhance model interpretability and credibility. Third, missing, frozen, and drifting data should be further studied to avoid the implicit amplification of data-quality problems within the model.

Acknowledgements

The authors would like to extend their sincere gratitude to North China University of Science and Technology in Tangshan, China, as well as to Petrobras for supplying the data used in this study.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Azmi, P.A.R., Yusoff, M. and Mohd Sallehud-din, M.T. (2024) A Review of Predictive Analytics Models in the Oil and Gas Industries. *Sensors*, **24**, Article 4013. <https://doi.org/10.3390/s24124013>
- [2] D'Almeida, A.L., Bergiante, N.C.R., de Souza Ferreira, G., Leta, F.R., de Campos Lima, C.B. and Lima, G.B.A. (2022) Digital Transformation: A Review on Artificial Intelligence Techniques in Drilling and Production Applications. *The International Journal of Advanced Manufacturing Technology*, **119**, 5553-5582. <https://doi.org/10.1007/s00170-021-08631-w>
- [3] Tariq, Z., Aljawad, M.S., Hasan, A., Murtaza, M., Mohammed, E., El-Husseiny, A., *et al.* (2021) A Systematic Review of Data Science and Machine Learning Applications to the Oil and Gas Industry. *Journal of Petroleum Exploration and Production Technology*, **11**, 4339-4374. <https://doi.org/10.1007/s13202-021-01302-2>
- [4] Vargas, R.E.V., Munaro, C.J., Ciarelli, P.M., Medeiros, A.G., Amaral, B.G.D., Barrionuevo, D.C., *et al.* (2019) A Realistic and Public Dataset with Rare Undesirable Real Events in Oil Wells. *Journal of Petroleum Science and Engineering*, **181**, Article ID: 106223. <https://doi.org/10.1016/j.petrol.2019.106223>
- [5] Marins, M.A., Barros, B.D., Santos, I.H., Barrionuevo, D.C., Vargas, R.E.V., de M. Prego, T., *et al.* (2021) Fault Detection and Classification in Oil Wells and Production/Service Lines Using Random Forest. *Journal of Petroleum Science and Engineering*, **197**, Article ID: 107879. <https://doi.org/10.1016/j.petrol.2020.107879>
- [6] Machado, A.P.F., Vargas, R.E.V., Ciarelli, P.M. and Munaro, C.J. (2022) Improving Performance of One-Class Classifiers Applied to Anomaly Detection in Oil Wells. *Journal of Petroleum Science and Engineering*, **218**, Article ID: 110983. <https://doi.org/10.1016/j.petrol.2022.110983>
- [7] Fernandes, W., Komati, K.S. and Assis de Souza Gazolli, K. (2023) Anomaly Detection in Oil-Producing Wells: A Comparative Study of One-Class Classifiers in a Multivariate Time Series Dataset. *Journal of Petroleum Exploration and Production Technology*, **14**, 343-363. <https://doi.org/10.1007/s13202-023-01710-6>
- [8] Bayazitova, G., Anastasiadou, M. and dos Santos, V.D. (2024) Oil and Gas Flow Anomaly Detection on Offshore Naturally Flowing Wells Using Deep Neural Networks. *Geoenery Science and Engineering*, **242**, Article ID: 213240. <https://doi.org/10.1016/j.geoen.2024.213240>
- [9] Soriano-Vargas, A., Werneck, R., Moura, R., Mendes Júnior, P., Prates, R., Castro, M., *et al.* (2021) A Visual Analytics Approach to Anomaly Detection in Hydrocarbon Reservoir Time Series Data. *Journal of Petroleum Science and Engineering*, **206**, Article ID: 108988. <https://doi.org/10.1016/j.petrol.2021.108988>
- [10] Lopes, L.G.O., Vieira, T.M.d.A., Aranha, P.E., Lima Junior, E.T.D. and Lira, W.W.M. (2025) Detection and Classification of Anomalies in Oil Well Production Using

- Open-World Learning. *Engineering Applications of Artificial Intelligence*, **159**, Article ID: 111514. <https://doi.org/10.1016/j.engappai.2025.111514>
- [11] Chandola, V., Banerjee, A. and Kumar, V. (2009) Anomaly Detection. *ACM Computing Surveys*, **41**, 1-58. <https://doi.org/10.1145/1541880.1541882>
- [12] Blázquez-García, A., Conde, A., Mori, U. and Lozano, J.A. (2021) A Review on Outlier/Anomaly Detection in Time Series Data. *ACM Computing Surveys*, **54**, 1-33. <https://doi.org/10.1145/3444690>
- [13] Breunig, M.M., Kriegel, H., Ng, R.T. and Sander, J. (2000) LOF: Identifying Density-based Local Outliers. *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, Dallas, 15-18 May 2000, 93-104. <https://doi.org/10.1145/342009.335388>
- [14] Liu, F.T., Ting, K.M. and Zhou, Z. (2008). Isolation Forest. 2008 *Eighth IEEE International Conference on Data Mining*, Pisa, 15-19 December 2008, 413-422. <https://doi.org/10.1109/icdm.2008.17>
- [15] Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J. and Williamson, R.C. (2001) Estimating the Support of a High-Dimensional Distribution. *Neural Computation*, **13**, 1443-1471. <https://doi.org/10.1162/089976601750264965>
- [16] Jardine, A.K.S., Lin, D. and Banjevic, D. (2006) A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance. *Mechanical Systems and Signal Processing*, **20**, 1483-1510. <https://doi.org/10.1016/j.ymssp.2005.09.012>
- [17] Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L. and Muller, P. (2019) Deep Learning for Time Series Classification: A Review. *Data Mining and Knowledge Discovery*, **33**, 917-963. <https://doi.org/10.1007/s10618-019-00619-1>
- [18] Malhotra, P., Vig, L., Shroff, G., et al. (2015) Long Short Term Memory Networks for Anomaly Detection in Time Series. *ESANN 2015 Proceedings*, Bruges, 22-24 April 2015, 89-94.
- [19] Hundman, K., Constantinou, V., Laporte, C., Colwell, I. and Soderstrom, T. (2018) Detecting Spacecraft Anomalies Using Lstms and Nonparametric Dynamic Thresholding. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, 19-23 August 2018, 387-395. <https://doi.org/10.1145/3219819.3219845>
- [20] Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W. and Pei, D. (2019) Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, 4-8 August 2019, 2828-2837. <https://doi.org/10.1145/3292500.3330672>
- [21] Li, D., Chen, D., Jin, B., Shi, L., Goh, J. and Ng, S. (2019) MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks. In: Tetko, I., Kůrková, V., Karpov, P. and Theis, F., Eds., *Artificial Neural Networks and Machine Learning—ICANN 2019: Text and Time Series*, Springer, 703-716. https://doi.org/10.1007/978-3-030-30490-4_56
- [22] Zong, B., Song, Q., Min, M.R., et al. (2018) Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection. 2018 *International Conference on Learning Representations*, Vancouver, 30 April-3 May 2018, 1-19.
- [23] Audibert, J., Michiardi, P., Guyard, F., Marti, S. and Zuluaga, M.A. (2020) USAD: Unsupervised Anomaly Detection on Multivariate Time Series. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 6-10 July 2020, 3395-3404. <https://doi.org/10.1145/3394486.3403392>
- [24] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need.

arXiv: 1706.03762

- [25] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., et al. (2021) Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 11106-11115. <https://doi.org/10.1609/aaai.v35i12.17325>
- [26] Xu, J., Wu, H., Wang, J., et al. (2022) Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy. arXiv: 2110.02642
- [27] Tuli, S., Casale, G. and Jennings, N.R. (2022) TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data. *Proceedings of the VLDB Endowment*, **15**, 1201-1214. <https://doi.org/10.14778/3514061.3514067>
- [28] Yang, Y., Zhang, C., Zhou, T., Wen, Q. and Sun, L. (2023) DCdetector: Dual Attention Contrastive Representation Learning for Time Series Anomaly Detection. *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Long Beach, 6-10 August 2023, 3033-3045. <https://doi.org/10.1145/3580305.3599295>
- [29] Zhao, M., Peng, H., Li, L. and Ren, Y. (2024) Graph Attention Network and Informer for Multivariate Time Series Anomaly Detection. *Sensors*, **24**, Article 1522. <https://doi.org/10.3390/s24051522>
- [30] Deng, A. and Hooi, B. (2021) Graph Neural Network-Based Anomaly Detection in Multivariate Time Series. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 4027-4035. <https://doi.org/10.1609/aaai.v35i5.16523>