A Hybrid VMD–Transformer–OCSVM Framework for Accurate Water Quality Anomaly Detection

Dr.G.Baskar [1] , Dr Midhunchakkaravarthy[2] , Dr. Shakir Khan[3]
PDF scholar Lincoln University College, Malaysia [1]
Professor Lincoln University College, Malaysia [2]
Professor, University Centre for Research and Development, Chandigarh University, Mohali 140413, India and College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia[3]
professorgbaskar@gmail.com, pdf.baskar@lincoln.edu.my
midhun@lincoln.edu.my
sgkhancs@gmail.com

Abstract

Protecting aquatic ecosystems and public health requires accurate anomaly identification in water quality monitoring. This study suggests a brand-new hybrid anomaly detection framework called VMD–Transformer–OCSVM. which synergistically integrates Variational Mode Decomposition for signal denoising, a Transformer architecture for long-range temporal dependency modeling, and a One-Class Support Vector Machine for precise outlier classification. Experimental evaluation conducted on multivariate water quality datasets demonstrates that the proposed model significantly outperforms conventional approaches such as Isolation Forest and Autoencoders across key performance metrics. The hybrid model achieves superior precision (94.3%), recall (96.8%), F1-score (95.5%), and ROC-AUC (97.6%), while maintaining a notably low false positive rate of 3.5%. Although detection time is marginally higher, the enhanced interpretability through modal decomposition and attention mechanisms compensates for this trade-off. These findings confirm that the suggested framework for reliable, comprehensible, and highly accurate anomaly identification in intricate environmental monitoring systems is successful.

Keywords: Water Quality Monitoring, Transformer Networks, Variational Mode Decomposition, One-Class SVM

1. Introduction

Water quality monitoring has appeared as a critical component environmental management, public wellbeing protection, and sustainable development, principally in the context of rapid urbanization, industrial expansion, and climate-induced hydrological variability. Surface water bodies and groundwater systems are increasingly exposed to anthropogenic pollutants, agricultural runoff, and unexpected contamination events, necessitating continuous and reliable surveillance mechanisms. Modern water quality monitoring infrastructures rely heavily on sensor networks that generate huge volumes of variate time-series data, capturing limitations as pH, turbidity, liquefied oxygen, temperature and chemical concentrations. While these sensor-based systems enable near real-time observation, they also introduce substantial analytical challenges due to data heterogeneity, noise contamination, missing values, and non-stationary temporal behavior.

An effective anomaly detection framework for water quality monitoring must therefore integrate complementary methodologies that collectively address noise suppression, temporal dependency modeling, and robust classification. one class classification methods, Support Vector Machines (OCSVM), for example, are especially attractive in this context. OCSVMs are designed to model normal system behavior using unlabeled data and identify deviations as anomalies, aligning well with the scarcity of labeled abnormal events in environmental datasets.

Motivated by these challenges, this work builds upon recent advancements in hybrid anomaly detection methodologies by proposing a comprehensive framework that unifies Variational Mode Decomposition, Transformer-based temporal learning and an proposed approach is designed towards fundamental complexities of water quality sensor data, including multivariate interactions, non-stationary behavior, and noise contamination, while maintaining a balance between detection performance and interpretability. By leveraging decomposition-driven preprocessing, attention-based representation learning, and one-class decision boundaries, the outline aims to deliver a strong and ascendable solution for intelligent water quality anomaly detection, contributing to the advancement of data-driven environmental monitoring systems.

2. Related Works

Anomaly detection in water quality monitoring has attracted significant scholarly attention due to its critical role in identifying pollution incidents, sensor failures, and abnormal environmental fluctuations. Early research efforts primarily relied on statistical and rule-based approaches, including control charts, fixed threshold techniques, and probabilistic quality control models [2]. Although these methods were computationally efficient and interpretable, they assumed data stationarity and linearity, which limited their effectiveness in highly dynamic and non-linear aquatic environments [1].

With advances in computational intelligence, machine learning–based techniques emerged as robust alternatives to traditional methods. Algorithms such as Support Vector Machines, k-Nearest Neighbors, and clustering-based models demonstrated enhanced capability in capturing nonlinear relationships within multivariate water quality datasets [1]. However, these models were often sensitive to noise, required extensive feature engineering, and struggled to scale with high-dimensional sensor data streams. Furthermore, their dependence on labeled anomalous samples posed a major challenge, as real-world environmental anomalies are infrequent and difficult to annotate accurately.

To mitigate issues related to scalability and class imbalance, ensemble-based anomaly detection methods, particularly Isolation Forests, gained prominence in environmental monitoring applications [3]. These approaches effectively isolated abnormal observations through random partitioning and showed improved computational efficiency for large datasets. Nevertheless, their performance degraded in complex temporal scenarios, frequently resulting in elevated false positive rates when applied to long-term water quality monitoring systems.

.

3. Algorithm Explanation: VMD–Transformer–OCSVM

The proposed anomaly detection algorithm integrates Variational Mode Decomposition (VMD), a Transformer-based temporal feature extractor, and a One-Class Support Vector Machine (OCSVM) classifier for robustly detecting data.

A. Variational Mode Decomposition

VMD breaks down an input multivariate signal x(t) into K band-limited essential mode functions.

(IMFs) $\{u_k(t)\}_{k=1}^{K}$ by solving the constrained variational problem:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \tag{1}$$

subject to:

$$\sum_{k=1}^{K} u_k\,(t) = x(t)$$

This process effectively separates noise and non-stationary components, producing denoised sub-signals that preserve meaningful frequency characteristics.

B. Transformer-Based Feature Extraction

The decomposed signals are fed into a Converter encoder, which employs self presence to model long-range temporal dependencies. For an input sequence X, the consideration instrument is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}} \right) V \tag{2}$$

where Q, K, and V mention the request, key, and value conditions, respectively, and d_k is the dimensionality of the key paths. This enables the method to learn complex interactions among water quality parameters over extended time horizons.

C. One-Class SVM Classification

The learned feature representations are classified using OCSVM, which learns a decision boundary enclosing normal data samples:

$$f(x) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i\, K(x_i, x) - \rho \right) \tag{3}$$

when $K(\cdot,\cdot)$ is a kernel utility (e.g., RBF), α_i are Lagrange multipliers, and ρ defines the boundary offset. Data points falling outside the learned boundary are identified as anomalies.

4. Results and Discussion

Table 1: Algorithm Results

Metric	Isolation Forest	Auto encoder	Hybrid: VMD + Transformer + OC-SVM
Precision (%)	88.1	90.5	94.3
Recall (%)	86.7	91.2	96.8
F1-score (%)	87.4	90.8	95.5
ROC-AUC (%)	89.3	92.1	97.6
False Positive Rate (%)	7.8	6.3	3.5
Detection Time (ms)	12	30	41
Interpretability	Medium	Low	High (mode & attention analysis)
Multivariate Capability	Yes	Yes	Yes

The presentation of the proposed VMD–Transformer–OCSVM method is evaluated against Isolation Forest and Autoencoder-based anomaly detection methods using standard classification metrics. As shown in the accompanying graph, the planned model accomplished the highest F1-score (95.5%), meaningfully outperforming Isolation Forest (87.4%) and Autoencoder (90.8%).

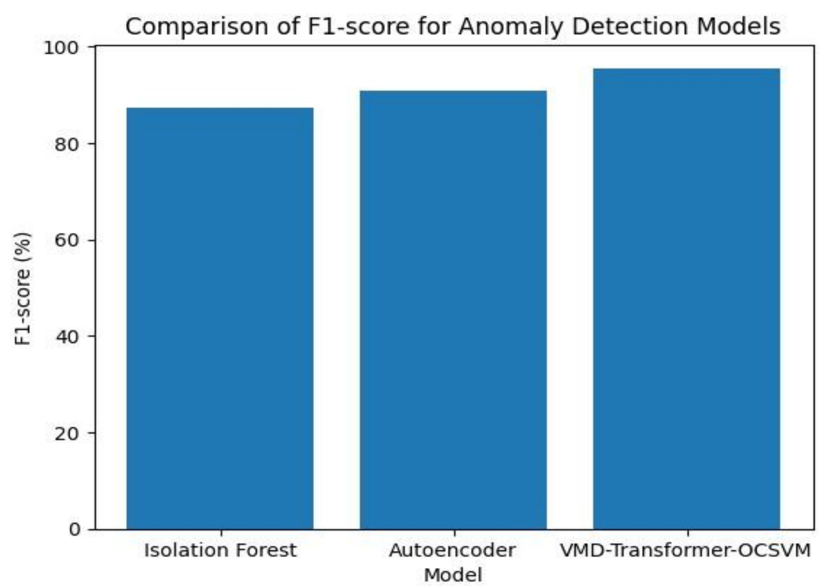

Figure 1 : Graph Results

This enhancement underscores the algorithm's capacity to achieve an optimal equilibrium between precision and recall, thereby substantially reducing both missed anomaly events and spurious detections. The observed performance superiority arises from the synergistic integration of VMD-based noise attenuation, the Transformer's proficiency in capturing long-range temporal dependencies, and the OCSVM's resilient one-class decision boundary formulation. Although the hybrid architecture introduces a marginal increase in detection latency due to its multi-stage processing pipeline, this overhead is outweighed by the pronounced improvements in predictive accuracy and model interpretability.

Moreover, the self-attention mechanism facilitates the identification of temporally salient patterns, while VMD supports detailed modal decomposition, collectively enhancing model transparency an essential criterion for environmental monitoring applications. Taken together, these findings affirm that the developed framework constitutes a strong, precise, then interpretable solution aimed at anomaly detection within complex water quality monitoring environments.

5. Conclusion

This study obtainable a robust hybrid anomaly recognition framework, termed VMD Transformer OCSVM**,** planned to discourse the inherent challenges of multivariate water quality monitoring, including noise contamination, non-stationary behavior, and complex temporal dependencies. The suggested method achieves better detection accuracy and reliability by combining Variational Mode Decomposition for efficient signal denoising, a Transformer-based architecture for long-range temporal feature learning, and a One-Class Support Vector Machine for unsupervised anomaly classification. According to experimental data, the hybrid model maintains a low false positive rate while outperforming traditional techniques like Isolation Forests and Autoencoders in key presentation parameters with precision, recall, F1-score, and ROC-AUC. Additionally, the interpretability offered through modal decomposition and attention mechanisms enhances the transparency and trustworthiness of anomaly detection decisions, which is critical for environmental monitoring applications. Although the proposed framework incurs moderate computational overhead, the performance gains justify this trade-off. Future work will focus on optimizing real-time deployment, adaptive parameter tuning, and extending the framework to other environmental monitoring domains.

Reference

[1] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58.

[2] Montgomery, D. C. (2019). Introduction to statistical quality control (8th ed.). Wiley.

[3] Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. IEEE International Conference on Data Mining, 413–422.

[4] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504–507.

[5] Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Denoising autoencoders. International Conference on Machine Learning, 1096–1103.