Graph-Based Adaptive Anomaly Detection Framework for Dual-Fuel Marine Engines

Jaewoong Choi[1], Yoojeong Noh[1], Young-Jin Kang[2]

[1]*Pusan National University, Busan, 46241, Republic of Korea*
jwnggggg@pusan.ac.kr
yoonoh@pusan.ac.kr

[2]*Pusan National University, EPIC, Busan, 46241, Republic of Korea*
zmanx@pusan.ac.kr

ABSTRACT

Dual-fuel (DF) marine engines, capable of operating on both diesel and LNG, face significant monitoring challenges due to frequent mode switching, dual valve timing, and load variability, which create nonlinear, time-varying dependencies among sensors. Such dynamics undermine conventional time-series anomaly detection methods that overlook structural relationships. To address this, we propose a graph-based anomaly detection framework tailored for DF engine monitoring. Sensor readings are modeled as nodes, with edges encoding domain-informed physical or functional dependencies. A multi-head Graph Attention Network (GAT)–based overcomplete autoencoder captures both local sensor behavior and global structural patterns; the expanded latent space preserves fine-grained features and heightens sensitivity to subtle deviations. The encoder aggregates context-aware features, and the decoder ensures graph-consistent reconstruction. Anomalies are scored using a λ-weighted combination of node-level reconstruction error (RMSE) and graph-level structural inconsistency from Graph Laplacian Smoothness (GLS). The λ parameter is optimized post hoc on validation data via F1-score, balancing sensitivity and precision. Evaluation on ten months of DF engine data demonstrates interpretable, real-time fault detection and sensor-level localization, supporting practical, condition-based maintenance.

1. INTRODUCTION

Recent developments in maritime environmental regulations, such as the International Maritime Organization's 2020 sulfur oxides (SOx) emission limits, have accelerated the adoption of dual-fuel (DF) marine engines that can operate on both diesel and liquefied natural gas (LNG) (IMO, 2020; Sigalas, 2022; Karatuğ et al., 2023). These engines not only enhance fuel efficiency and reduce exhaust emissions, but also expand the potential for implementing condition-based maintenance (CBM) and voyage optimization strategies (Mohamad et al., 2021; Sutrisno et al., 2025). However, frequent switching between operating modes, dual valve timing, and fluctuating loads introduce considerable operational complexity, resulting in highly nonlinear and time-varying interdependencies among onboard sensors (Youssef et al., 2024; Elahi et al., 2023). Consequently, conventional time-series-based anomaly detection methods often fail to capture the underlying structural correlations between sensors, leading to frequent false alarms and missed faults under dynamic operating conditions (Iqbal et al., 2024).

To address these challenges, recent research has explored graph-based deep learning methods that model the sensor network as a graph, where nodes represent individual sensors and edges denote functional or physical dependencies, thereby enabling the representation of the system's structural topology (Veličković et al., 2018; Zhao et al., 2024). In particular, graph attention networks (GATs) dynamically learn attention coefficients to capture complex and heterogeneous sensor interactions, achieving superior performance compared to conventional CNN- or RNN-based approaches under highly variable operating conditions (Ding et al., 2023).

Nevertheless, existing GAT-based anomaly detection studies exhibit two major limitations. First, most approaches rely solely on node-level reconstruction errors or classification losses and do not explicitly account for global graph structural consistency. As a result, the ability to detect distributed anomalies—subtle perturbations to the dependency structure among sensors—is limited. Second, the weighting parameter (λ) for combining multiple anomaly indicators is seldom optimized to align with practical CBM objectives, which prevents achieving an optimal balance between detection sensitivity and false alarm rate. These limitations significantly hinder the applicability of such methods in real-world maritime CBM implementations.

To overcome these limitations, we propose an unsupervised anomaly detection framework that integrates a GAT-based encoder with an overcomplete autoencoder architecture, enabling the simultaneous capture of both local and global anomalies. The GAT-based encoder learns context-aware node embeddings by incorporating each sensor's local operational behavior together with the global structural dependencies of the network. The decoder then reconstructs the original graph signals while preserving these structural relationships.

During the reconstruction process, two complementary anomaly indicators are computed. The first is the node-level reconstruction error, quantified by the mean squared error (MSE), which is sensitive to local anomalies. The second is the graph Laplacian smoothness (GLS), which measures the global structural consistency of the graph and reflects disruptions in physical or functional dependencies across the sensor network.

A composite anomaly score is derived by calculating a weighted sum of these two indicators, with λ denoting the weighting factor. This parameter is post hoc optimized on the validation set to maximize the F1-score, thereby achieving a balanced trade-off between early fault detection sensitivity and the reduction of false alarm rates (Gharib & Kovacs, 2024; Sun et al., 2024). Furthermore, per-node GLS scores provide interpretable localization of anomalies at the sensor level, while a mimic board-based visualization facilitates operational decision-making and CBM prioritization in the complex engine room environment (Young et al., 2023; Jovanović et al., 2025).

The proposed framework was validated using a 10-month dataset collected from an operational DF marine engine. Experimental results demonstrate that the method achieves both high detection accuracy and interpretability, effectively overcoming the inherent weaknesses of prior approaches. This establishes a robust and scalable foundation for CBM implementation in marine propulsion systems operating under variable and challenging maritime conditions.

2. OVERVIEW OF THE PROPOSED METHOD

The proposed method is a graph-based anomaly detection and fault diagnosis framework designed to address the complex operating conditions of dual-fuel (DF) marine engines. As illustrated in Fig. 1, the process begins with the preprocessing of high-resolution operational data collected from 35 key onboard sensors, recorded during normal operating conditions. Each measurement at a given timestamp is modeled as a static directed graph, where each node represents a sensor and edges encode physical or functional dependencies derived from domain expertise and system schematics. The resulting engine-system graph is fed into a multi-head Graph Attention Network (GAT)-based overcomplete autoencoder, which employs attention mechanisms to capture both localized sensor behavior and global structural interactions. The encoder maps node features into a higher-dimensional latent space to preserve fine-grained patterns, while the decoder reconstructs the input feature matrix in a manner consistent with the original graph topology. Anomalies are quantified via a λ-weighted composite anomaly score, integrating node-level reconstruction error (root mean square error, RMSE) and graph-level structural inconsistency measured by Graph Laplacian Smoothness (GLS). The λ parameter is not fixed a priori; instead, it is post hoc optimized on validation data to

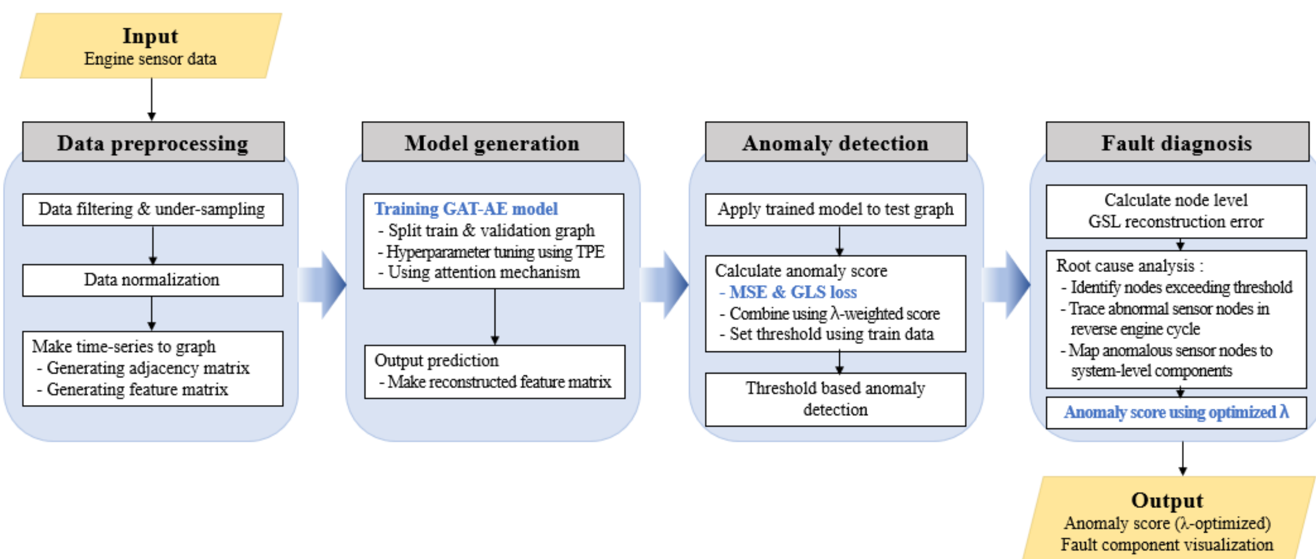

Figure 1. Overview of the proposed method

maximize the F1-score, enabling an adaptive trade-off between sensitivity and precision according to operational priorities. Finally, node-wise GLS reconstruction errors are analyzed to localize the sensors contributing most to structural anomalies, and the results are visualized on a sensor-level mimic board. This end-to-end process supports both early fault detection and interpretable diagnosis, making it well-suited for practical condition-based maintenance (CBM) in real-world DF marine engine operations.

3. DATA DESCRIPTION

The target system in this study is a dual‐fuel (DF) marine engine that can operate on both diesel and liquefied natural gas (LNG). The engine is equipped with low‐emission combustion technology, dual valve timing (DVT), and a turbocharger control mechanism, which result in nonlinear and time‐varying interactions between onboard sensors under diverse operating conditions such as fuel‐mode switching, load variation, and valve timing adjustments.

The dataset was collected over an extended period during actual operation of a vessel equipped with a DF engine. It contains hundreds of thousands of time‐series entries recorded at one‐minute intervals, with measurements from more than thirty key sensors. These sensors are grouped into six subsystems: engine control, combustion/air system, gas system, diesel system, cylinder system, and mechanical system. Representative variables include engine load, fuel mode, charge air pressure, peak cylinder pressure (P-max), exhaust gas temperature, fuel oil pressure, gas pressure, turbocharger speed, and main bearing temperature. This sensor configuration was designed to comprehensively represent the DF engine's fuel usage patterns, combustion characteristics, operating conditions, and mechanical health.

Data preprocessing was conducted to improve overall data quality, address imbalance among sensor types and conditions, and prepare inputs for subsequent graph‐based modeling. First, non‐representative intervals, including idle runs and abnormal operating ranges, were removed so that only records corresponding to normal load operation were retained. Extreme outliers caused by sensor faults, communication errors, or power instabilities—values physically implausible or outside the normal operating bounds—were also excluded. No interpolation or imputation was applied; only trustworthy data were used for analysis. To mitigate scale bias between variables, a two‐stage normalization was applied: a Robust Scaler to reduce the impact of outliers using median and interquartile range, followed by Min–Max scaling to normalize values to the range (Choi et al. 2025).

Additionally, undersampling was applied to certain overrepresented load–fuel mode combinations in order to balance the dataset and ensure wider coverage of operational conditions. The resulting refined and balanced dataset served as a reliable basis for training the proposed graph‐based anomaly detection model, supporting both model stability during training and improved generalization performance.

4. GRAPH-BASED FAULT DETECTION AND DIAGNOSIS

4.1. Model Development and Training

The GAT employed in this study processes each timestamp by modeling the preprocessed sensor measurements as a static directed graph $G = (V, E, X)$. Here, nodes V correspond to sensors, edges E encode physical and functional dependencies based on domain expertise, and $X \in \mathbb{R}^{N\times F}$ is the node feature matrix with N sensors and F features per node. This representation facilitates the capture of both local sensor behavior and global structural patterns.

The GAT encoder computes attention coefficients between a target node i and each neighbor $j \in N(i)$ using

$$e_{ij} = LeakyReLU(a^{\mathrm{T}}[Wh_{\mathrm{i}} \parallel Wh_{\mathrm{j}}]) \tag{1}$$

where W is a learnable weight matrix, h_i and h_j are transformed node feature vectors, a is the attention weight vector, and || denotes concatenation.

These coefficients are normalized by the softmax function to give

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{\{k\in\mathcal{N}_{\mathrm{i}}\}} exp(e_{ik})} \tag{2}$$

where α_{ij} expresses the relative importance of neighbor j to node i . By adopting an overcomplete latent space (dim(Z)>dim(X)), the encoder preserves fine-grained sensor patterns, while the decoder reconstructs the node feature matrix $\hat{\boldsymbol{X}}$ with graph topology consistency.

Training is conducted in an unsupervised manner using only normal-operation data. The reconstruction objective minimizes the mean squared error (MSE) between input and output:

$$MSE = \frac{1}{N}\sum_{i=1}^{N} \|x_i - \hat{x}_i\|^2 \tag{3}$$

where $x_i \in \mathbb{R}^F$ is the original feature vector of node i and $\hat{x}_i$ is its reconstruction.

Training is performed with mini-batch graph inputs using the PyTorch Geometric framework, and the key hyperparameters such as hidden dimension size, number of attention heads, learning rate, and dropout rate are tuned by Tree-structured Parzen Estimator(TPE)–based Bayesian optimization. This ensures robust generalization and prevents overfitting.

4.2. λ Post-hoc Optimization

Anomalies are quantified by a composite anomaly score that integrates two complementary loss components: (i) node-level reconstruction error (RMSE) and (ii) graph-level structural inconsistency measured as the difference in Graph Laplacian Smoothness (GLS) between input and reconstructed graphs. GLF

$$S(X) = Tr(X^T LX) \tag{4}$$

where $L = I - D^{-1/2}AD^{-1/2}$ is the normalized Laplacian (A: adjacency matrix, D: degree matrix, I: Identity), and $Tr(\cdot)$ denotes the matrix trace. The structural loss is the difference in smoothness before and after reconstruction:

$$L_{graph} = \left|S(X) - S(\hat{X})\right| \tag{5}$$

The normalized node-level and graph-level losses are combined as

$$A_{score} = \lambda \cdot MSE_{norm} + (1 - \lambda) \cdot L_{graph,norm} \tag{6}$$

where the weighting coefficient λ adjusts the relative contribution of the two terms. Lower λ values emphasize early detection of structural anomalies, while higher values prioritize precision in identifying localized sensor faults. Because anomaly detection in real-world DF engine data often involves a strong class imbalance, where accuracy alone can be misleading, the F1-score was adopted to jointly account for both precision and recall. Instead of fixing λ heuristically, a post-hoc optimization procedure was applied: a grid search over candidate λ values was conducted on labeled test data to compute precision, recall, and F1-score, with the λ maximizing the F1-score selected as optimal. This optimization enables flexible tuning of the sensitivity–precision trade-off to match operational requirements for either early warning or false alarm suppression.

$$\lambda^* = \underset{\lambda}{\operatorname{argmax}}\, F1 - score(\lambda) \tag{7}$$

4.3. Results and Discussion

For interpretability and fault localization, node-wise Graph Laplacian Smoothness (GLS) reconstruction errors were analyzed for each detected anomaly. These values quantify the degree to which the structural relationships of a given node with its neighbors degrade during reconstruction, directly indicating the node's contribution to the anomaly. Nodes were classified as suspicious (above the 95th percentile) or anomalous (above the 99th percentile) based on dual thresholds derived from the training data distribution.

The proposed GAT-based overcomplete autoencoder model was applied to real operational data from a dual-fuel (DF) marine engine to evaluate the impact of the weighting parameter λ in the integrated anomaly score. For a representative abnormal event, the node-level reconstruction error (MSE), graph-level structural loss (GLS), and their λ-weighted combination were analyzed, and the detection results are summarized in Table 1.

Figure 2 compares the detection results for main bearing temperature anomalies against engine load for λ values of 0.1, 0.567, and 0.8. Detected anomalies are color-coded—red for true positives (T.P), blue for false negatives (F.N), and orange for false positives (F.P)—while normal operating points are displayed in blue. When λ = 0.1, the GLS component dominates, making the model highly sensitive to subtle structural changes but increasing false positives. When λ = 0.8, the MSE component dominates, focusing on sharp sensor-specific deviations but missing anomalies that mainly manifest as structural disruptions. At the optimal λ = 0.567, the model maintains balanced performance, reducing false positives while ensuring timely detection.

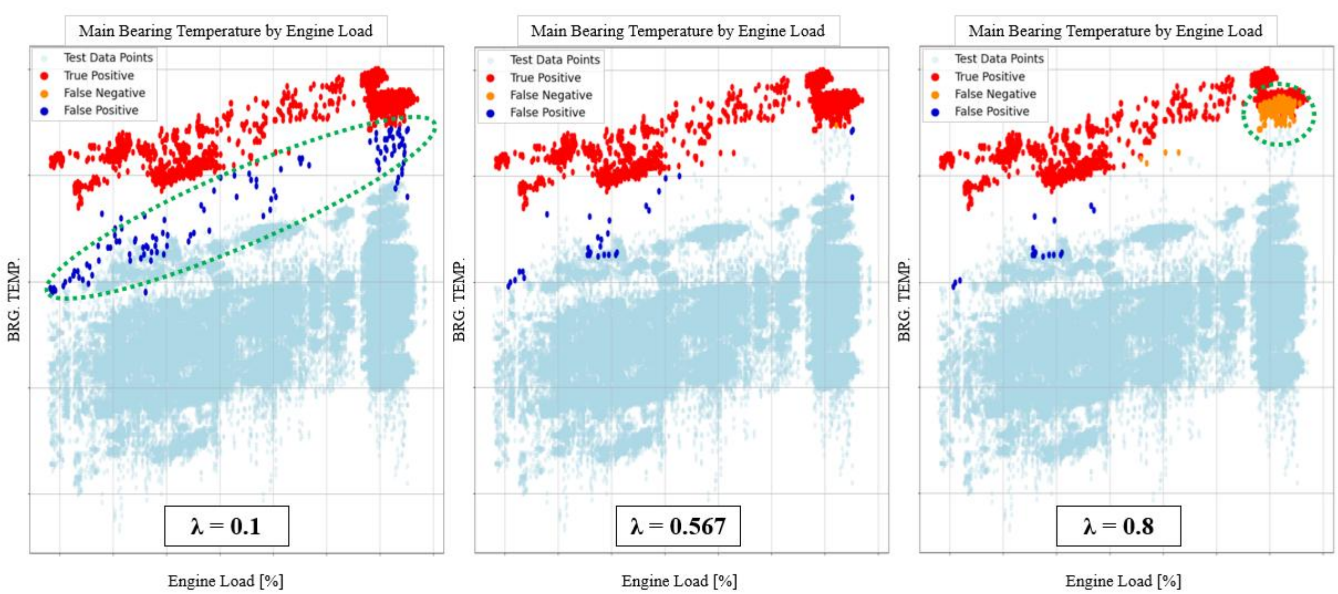

Figure 2. Anomaly detection results for different λ values

Overall, λ serves as a flexible tuning parameter that shifts the detection focus between localized and structural anomalies. When combined with GLS-based node-level diagnostics and intuitive visualizations such as the mimic board, the framework delivers both early and accurate anomaly detection with explanatory context, making it suitable for practical condition-based maintenance of DF marine engines.

Table 1. Effect of values on anomaly score composition

	λ = 0.1	λ = 0.567	λ = 0.8
MSE Loss	0.5425		
GLS Loss	0.7670		
MSE Term in Score	0.0543	0.3125	0.4340
GLS Term in Score	0.6903	0.3321	0.1534
Anomaly Score	0.7446	0.6446	0.5874
Threshold	0.6132		
Detection Result	Abnormal	Abnormal	Normal
Final Result	T.P	T.P	F.N

5. CONCLUSION

In this study, we proposed a Graph Attention Network (GAT)-based autoencoder framework to enable effective anomaly detection and fault diagnosis in dual-fuel (DF) marine engines operating under complex structural and dynamic conditions. Addressing the limitations of conventional time-series-based methods, which often fail to account for inter-sensor structural dependencies, the proposed approach models physical and functional interactions among sensors as a fixed graph derived from domain knowledge.

The framework integrates node-level reconstruction error and graph-level structural loss into a unified anomaly score, enabling the detection of both localized sensor faults and global structural anomalies. The weighting factor λ between the two loss terms is not manually set; instead, it is post hoc optimized to maximize the F1-score on a validation set. This data-driven optimization balances early fault detection with false alarm reduction, allowing the detection sensitivity to be dynamically tailored to operational requirements.

Experimental validation on ten months of real DF engine operational data confirmed that the model delivers high accuracy across diverse fault scenarios. Notably, it successfully detected subtle topological changes in the sensor network preceding main bearing overheating events, thereby providing earlier warnings than single-metric baselines. Furthermore, the use of mimic-board visualizations enabled intuitive differentiation between localized and structural anomalies, improving interpretability and supporting informed maintenance decisions.

In conclusion, the proposed GAT-based overcomplete autoencoder offers a robust, interpretable, and scalable framework for CBM in DF marine engines. With its demonstrated ability to adaptively balance sensitivity and precision, it provides a strong foundation for future applications in autonomous ship engine room operations. Future work will extend the framework to multi-engine graph modeling, leverage Shop test and Sea trial pretraining, and integrate reinforcement learning for fully autonomous fault diagnosis and End-to-End anomaly response.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2020-NR049569 and RS-2024-00341872).

REFERENCES

Yong, B. X., & Brintrup, A. (2022). Do autoencoders need a bottleneck for anomaly detection? IEEE Access, 10, 78455-78471.

Choi, J. (2025). Fault diagnosis of dual-fuel marine engines using graph neural networks [Master's thesis, Pusan National University].

Ding, C., Sun, S., & Zhao, J. (2023). MST-GAT: A multimodal spatial–temporal graph attention network for time series anomaly detection. *Information Fusion, 89*, 527–536.

Dong, X., Thanou, D., Frossard, P., & Vandergheynst, P. (2016). Learning Laplacian Matrix in Smooth Graph Signal Representations. *IEEE Transactions on Signal Processing*, 64(23), 6160–6173.

Elahi, M., Afolaranmi, S. O., Martinez Lastra, J. L., & Perez Garcia, J. A. (2023). A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment. *Discover Artificial Intelligence*, 3(1), Article 43.

Stoumpos, S., & Theotokatos, G. (2021). A novel methodology for marine dual fuel engines sensors diagnostics and health management. International Journal of Engine Research, 23(6), 974–994.

IMO. (2020). International Convention for the Prevention of Pollution from Ships (MARPOL) Annex VI.

Iqbal, A., Amin, R., Alsubaei, F. S., & Alzahrani, A. (2024). Anomaly detection in multivariate time series data using deep ensemble models. *PLOS ONE*, 19(6), e0303890.

Karatuğ, C., Ejder, E., Tadros, M., & Arslanoğlu, Y. (2023). Environmental and economic evaluation of dual-fuel engine investment of a container ship. *Journal of Marine Science and Application*, 22, 823–836.

Kong, Q., Chiang, A., Aguiar, A. C., Fernández-Godino, M. G., Myers, S. C., & Lucas, D. D. (2021). Deep Convolutional Autoencoders as Generic Feature Extractors in Seismological Applications. arXiv preprint arXiv:2110.11802.

Mohamad, F. F., Semin, B. C., Abu, B. R., & Birouk, M. (2021). Performance and emissions of natural gas/diesel dual-fuel engine at low load conditions: Effect of natural gas split injection strategy. *Fuel*, 300, 121012.

Sigalas, C. (2022). Financial impact of the IMO 2020 regulation on dry bulk shipping. *Maritime Transport Research*, 3(2), 100064.

Sutrisno, I., Umasangadji, F., Ardiansyah, A., Almuzani, N., Bashori, A., Mudjiono, U., Priyonggo, P., & Purwanti, E. P. (2025). Development of an IoT-Based Ship Engine Performance Monitoring System to Enhance Operational Efficiency. *Formosa Journal of Computer and Information Science*, 4(1), 83–92.

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio', P., & Bengio, Y. (2017). Graph Attention Networks. ArXiv, abs/1710.10903.

Verdoja, F., & Grangetto, M. (2020). Graph Laplacian for image anomaly detection. *Machine Vision and Applications*, 31(6), 1–12.

Youssef, A., Noura, H., El Amrani, A., El Adel, E. M., & Ouladsine, M. (2024). A survey on data-driven fault diagnostic techniques for marine diesel engines. IFAC-PapersOnLine, 58(4), 55–60.

Kang, Y.-K. Noh, Y., Jang, M., Park, S., Kim, J.-T., Kim, K.C. (2023). Hierarchical level fault detection and diagnosis of ship engine systems. *Expert Systems with Applications*, 213, 118814.

Zhao, M., Peng, H., Li, L., & Ren, Y. (2024). Graph Attention Network and Informer for Multivariate Time Series Anomaly Detection. *Sensors*, 24(5), 1522.