I. INTRODUCTION

Gas pipeline infrastructure plays a crucial role in ensuring the supply of essential energy for electricity and industrial needs worldwide [1]. However, its safety and operational reliability are often threatened by various risks and hazards, including the potential occurrence of anomalies or disruptions in operational data. These anomalies can arise from factors such as leaks that compromise safety, infrastructure damage that disrupts overall system performance, or operational inefficiencies that lead to economic losses [2]. Given the critical role of pipelines in energy distribution, early detection and rapid response to anomalies are essential to mitigate risks, prevent accidents, and ensure uninterrupted operation.

One promising approach to addressing these challenges is anomaly detection using machine learning. This approach involves programming machines to identify and flag data deemed unusual or suspicious. Unlike supervised learning, which relies on labeled data for training, anomaly detection methods can operate without labels. These models identify anomalies based on reconstruction errors, deviations from expected patterns, or statistical irregularities in the data [3]. The ability to function in unlabeled environments makes these methods highly suitable for complex datasets, such as timeseries data, often encountered in industrial systems like oil and gas pipelines.

Various anomaly detection techniques have been developed and successfully applied in different domains. Statistical methods like ARMA [4] and ARIMA [5] focus on modeling timeseries data trends and seasonal variations. Classical machine learning approaches such as PCA, k-Means, and KNN [6], [7], [8] provide baseline anomaly detection capabilities, while advanced deep learning methods, including autoencoders, variational autoencoders (VAE), and LSTM [3], [9], [10], have demonstrated remarkable performance in handling large-scale multivariate datasets. Hybrid approaches like VAE-LSTM [11] and GAN-based models [12] further enhance detection accuracy, especially in scenarios involving complex data patterns.

In domain-specific research, Ihsan and Astuti [3] leveraged deep learning autoencoders to reconstruct operational data from gas pipelines, enabling anomaly detection through reconstruction errors. Similarly, Aljameel et al. [2] applied machine learning models like SVM and random forests to detect leaks in oil and gas pipelines, achieving a 97.4% accuracy rate. Fernandes et al. [13] explored one-class classifiers, while Aranha and Lopes [14] combined rule-based systems with LSTM autoencoders to monitor oil well production anomalies. Recent advancements, such as the Digital Twin framework with MTAD-GAN proposed by Lian et al. [15], integrate virtual-real synchronization and attention mechanisms to detect anomalies in multivariate time-series data effectively.

This paper focuses on the application of Classic Seasonal Decomposition and Level Shift Anomaly Detection to detect anomalies in gas pipeline operational data, with the integration of deep learning models such as LSTM and VAE-GAN. Classic Seasonal Decomposition is used to separate time series data into trend, seasonal, and residual components, which provide deep insights into temporal patterns, including periodic fluctuations and long-term variations [16]. The resulting residual components allow the identification of unexpected patterns or anomalies that are not clearly visible in the original data.

As a next step, Level Shift Anomaly Detection is applied to detect sudden changes in the average data value, such as significant spikes or drops. These changes often indicate critical conditions, such as system failures, operational disruptions, or significant environmental changes [17]. This method is also used to label the dataset, which then becomes the basis for evaluating the performance of deep learning approaches such as LSTM [18] and VAE-GAN [19]. The combination of these two classical and modern methods aims to create a more accurate and reliable anomaly detection framework in managing time-based sequential data in gas pipeline systems.

*: Corresponding Authors