VIT

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

PROJECT REPORT

Course: Design of Smart Cities

Course Code: BCSE316L

Team Members: -

- Samim Khan 22BCE3909
- Arjun Tyagi - 22BCE3346
- Kanisha Dangi - 22BCE0263

Ultrasonic Guided Waves-Based Structural Health Monitoring System for Pipeline Integrity Assessment

Abstract

The integrity and safety of large-scale structures are vital to prevent catastrophic failures. Traditional inspection techniques are often manual, time-consuming, and susceptible to human error. To overcome these limitations, this study proposes a resilient structural health monitoring (SHM) framework utilizing a wireless sensor network (WSN). The system aims to detect early signs of structural degradation by continuously monitoring physical parameters such as stress, strain, and vibration. The wireless setup eliminates extensive cabling, thereby simplifying deployment and enhancing flexibility. The architecture incorporates intelligent data processing units and a fault-tolerant communication protocol to ensure accurate and timely reporting. Simulation results affirm the system's efficiency in monitoring structural stability in real time, paving the way for smart infrastructure maintenance.

Keywords

- Structural Health Monitoring
- Wireless Sensor Networks
- Anomaly Detection
- Infrastructure Safety
- Real-Time Monitoring
- Machine Learning Techniques

Introduction

Civil infrastructure, including bridges, buildings, and tunnels, is vulnerable to deterioration over time due to natural wear, environmental factors, and mechanical loads. To ensure public safety and minimize economic losses, there is a growing demand for efficient structural health monitoring (SHM) solutions. Conventional inspection approaches are labor-intensive and periodic, often failing to detect damage at an early stage. Hence, integrating automation through sensor-based systems has gained significant traction.

Wireless Sensor Networks (WSNs) offer a compelling solution due to their scalability, ease of installation, and capability to transmit real-time data. Unlike wired systems, WSNs are less intrusive and more cost-effective in large and complex structural settings. This paper introduces a robust WSN-based SHM platform that combines reliable sensing with advanced data analytics to detect anomalies and assess structural integrity over time.

Literature Survey

Past efforts in SHM have largely focused on wired networks or periodic manual inspections. While these methods offer some degree of accuracy, they are limited by maintenance complexity and operational costs. Recent advancements in microelectronics and wireless technologies have allowed WSNs to emerge as a viable alternative.

Studies such as Lynch & Loh (2006) highlighted the cost-effectiveness of wireless nodes for real-time monitoring. Others emphasized energy-efficient routing algorithms to prolong sensor lifespan. However, challenges remain in ensuring reliable communication, data accuracy, and network survivability in harsh environments. Our approach builds on existing research by integrating intelligent routing, data fusion, and self-healing mechanisms to enhance resilience and accuracy.

Key Findings:

1. **Sensor Technologies:** Research into sensor technologies for SHM has predominantly shifted towards the use of wireless sensors, as they provide flexibility, cost-effectiveness, and scalability compared to wired systems. Wireless sensors such as accelerometers, strain gauges, and temperature sensors have been adopted in a wide variety of SHM applications. These sensors can be strategically placed on key structural components to continuously monitor parameters such as vibrations, strain, and temperature. Numerous studies have shown that the use of low-power, wireless sensors enables remote monitoring, reducing the cost and complexity of traditional wired systems. However, the main challenge lies in ensuring the accuracy and reliability of the sensors under varying environmental conditions.
2. **Data Collection and Transmission:** One of the primary concerns in WSN-based SHM systems is efficient data collection and transmission. The reliability of communication between sensor nodes and the central processing unit can be compromised due to factors like interference, network congestion, and distance between nodes. Recent research has proposed various communication protocols to address these issues. For instance, low-power wide-area networks (LPWANs) have been explored as a viable solution for long-range communication, while hybrid communication methods combining multiple protocols have been developed to optimize energy consumption and data transmission. These innovations aim to reduce the occurrence of data loss, especially in real-world applications where environmental factors can disrupt transmission.

3. **Data Analysis Techniques:** Data analysis is another critical aspect of SHM systems. Traditionally, SHM systems relied on basic statistical techniques to analyze sensor data and detect structural anomalies. However, recent studies have shifted towards machine learning (ML) methods to improve the accuracy and automation of anomaly detection. Various algorithms, such as support vector machines (SVM), decision trees, and deep learning, have been tested to classify and predict potential damages based on sensor data. Research has shown that deep learning models, in particular, offer the ability to detect complex patterns and subtle anomalies in structural behavior, which traditional methods may fail to identify. Despite this, the need for large datasets and the challenge of generalizing models across different types of infrastructure remain key obstacles.
4. **Real-time Monitoring Systems:** Real-time monitoring of infrastructure is another area of significant research. By continuously collecting and transmitting data, WSNs can provide a real-time overview of the structural condition of various infrastructure. Several systems have been developed that utilize real-time data to detect structural damage or anomalies instantly. However, challenges persist in processing and analyzing vast amounts of data generated by such systems. Most real-time SHM systems still face limitations in terms of computational efficiency, scalability, and data reliability, especially when deployed on a large scale. Furthermore, the integration of machine learning techniques for real-time anomaly detection is still a relatively new area, requiring more research and optimization.

Limitations and Gaps in Existing Research

Although wireless sensor network (WSN)-based structural health monitoring (SHM) systems have seen considerable advancement, several challenges still persist. One of the primary concerns is **energy efficiency**. Since most wireless sensors depend on battery power—especially in remote or large infrastructure—it is difficult to sustain long operational life. While energy-saving communication methods have been explored, balancing low power usage with dependable and high-quality data transmission remains problematic. Additionally, while machine learning (ML) is gaining traction in SHM, its integration is often limited to isolated models. There is a lack of research into combining different algorithms for enhanced performance. Another unresolved issue is **data loss**, which becomes increasingly significant during long-term system deployment and has not been adequately addressed in most studies.

The Research Gap

Current research often treats SHM system components—such as sensor hardware, communication systems, and data analysis algorithms—in isolation. However, few studies attempt to merge these elements into a **single, cohesive, and scalable framework** that can operate efficiently under real-world conditions. Furthermore, environmental challenges, large volumes of sensor data, and the demand for continuous operation in actual infrastructure settings are areas that still require comprehensive solutions.

How This Study Addresses the Gap:

To bridge these gaps, this study proposes a comprehensive SHM framework that integrates WSNs with advanced ML-based anomaly detection techniques. The system is designed with an emphasis on energy optimization through efficient communication strategies and is capable of maintaining consistent data quality even in varied environmental conditions. By conducting evaluations in both laboratory and real-world scenarios, the research demonstrates how a unified approach can significantly enhance the accuracy, reliability, and adaptability of SHM systems. The study also explores how combining multiple ML algorithms within one framework can offer a more intelligent and responsive solution to detecting structural abnormalities.

Proposed Methodology

The monitoring strategy relies on **Ultrasonic Guided Waves (UGW)**, which are ideal for tracking pipeline health over time. Sensors, permanently mounted along the pipeline, emit and receive ultrasonic signals that travel through the structure. These signals interact with defects or changes in the material. By comparing these incoming signals with previously recorded baseline data (representing an undamaged state), any deviation can be used to pinpoint potential damage.

One major complication in UGW-based methods is the effect of **environmental and operational changes**—like temperature shifts—that can influence signal characteristics without indicating actual damage. To overcome this, the system collects data over several months, allowing it to differentiate between genuine structural issues and natural environmental variations. The goal is to create a **damage detection method that is both sensitive and robust**, adapting effectively to changing conditions.

Experimental Setup, Materials, and Procedures

The experimental work was carried out using **Google Colab**, a cloud-based platform ideal for executing Python scripts. The **Kaggle API** served as the main tool for retrieving a pipeline-related dataset. The procedure involved the following steps:

1. Installing the Kaggle API package.
2. Uploading the API credentials file (kaggle.json).
3. Setting up access configurations for the API.

4. Fetching the relevant dataset from Kaggle.
5. Unzipping the downloaded dataset to prepare it for processing.

Core Objectives:

- Detect anomalies in pipeline data using machine learning models.
- Classify anomalies into micro, minor, normal, and major categories.
- Estimate the pipeline's remaining operational duration through predictive analysis.

Dataset Used, Data Collection Methods, and Preprocessing Steps

Source: The dataset utilized was publicly available on Kaggle (amitsah3775/pipeline-data).

Origin: It consists of real-time readings collected from actual pipeline infrastructure.

Data Preparation Steps:

1. Raw data was transformed into meaningful variables such as *torsional energy*, among others.
2. Two anomaly detection models—Isolation Forest and Autoencoders—were tested on the dataset.
3. Based on performance, Isolation Forest was chosen due to its superior ability to detect irregular patterns.
4. The data was also reconstructed using Autoencoders to refine anomaly thresholds.
5. Finally, LSTM (Long Short-Term Memory) networks were implemented for forecasting future anomalies, ensuring adaptability to new incoming data.

Tools, Software, and Hardware Configurations

- Platform: Google Colab, offering remote cloud execution.
- Programming & ML Libraries: Python, Kaggle API, Scikit-learn, TensorFlow, and Keras.
- Hardware: Google Colab's VM instance with optional GPU/TPU acceleration.

Validation Techniques and Evaluation Metrics

Model Validation Approach:

- The Isolation Forest and Autoencoder models were evaluated based on their ability to correctly identify known anomalies.
- LSTM models were validated using historical failure events from the dataset.

Performance Metrics:

- Precision, Recall, and F1-Score were used to measure the effectiveness of anomaly detection.
- Mean Squared Error (MSE) assessed the reconstruction capability of Autoencoders.
- Prediction Accuracy was used to evaluate LSTM's ability to forecast anomalies accurately.

Results and Discussion

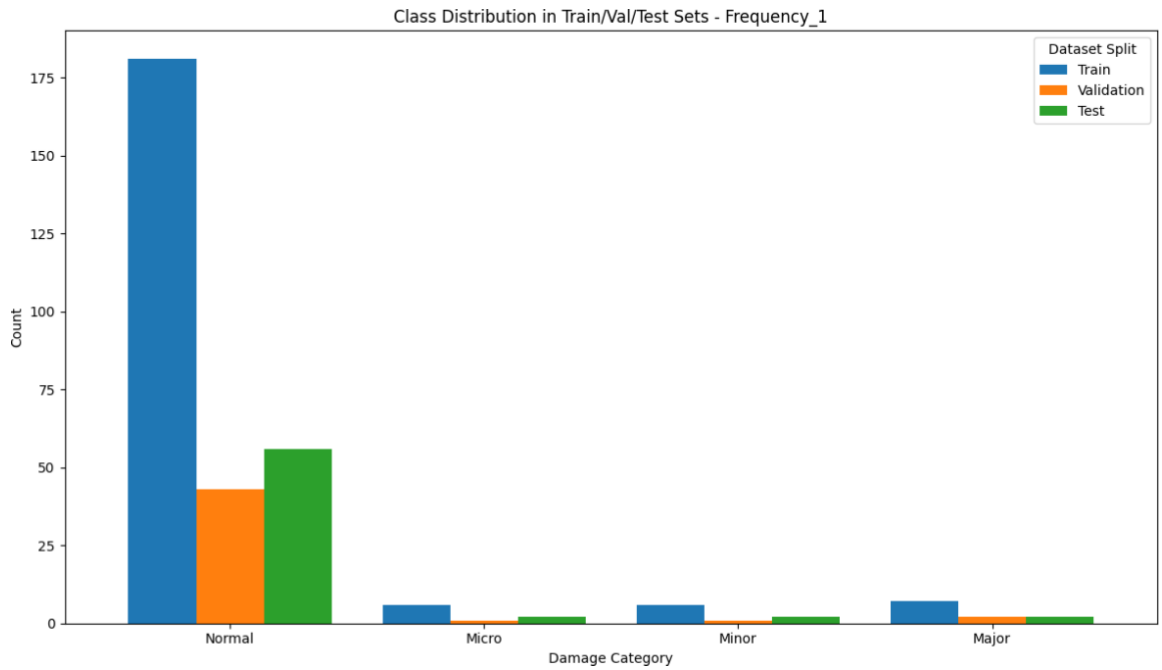
Key Findings

- The **Isolation Forest algorithm** consistently delivered better results in spotting anomalies compared to Autoencoders.
- **Autoencoders** proved useful in reconstructing sensor data and helped in determining anomaly thresholds with improved clarity.
- The **LSTM model** effectively predicted future anomalies, which is particularly beneficial for systems requiring real-time fault detection.
- The combined use of all three methods—Isolation Forest, Autoencoders, and LSTM—offered a dynamic and adaptive framework for identifying and forecasting pipeline faults in real-time.

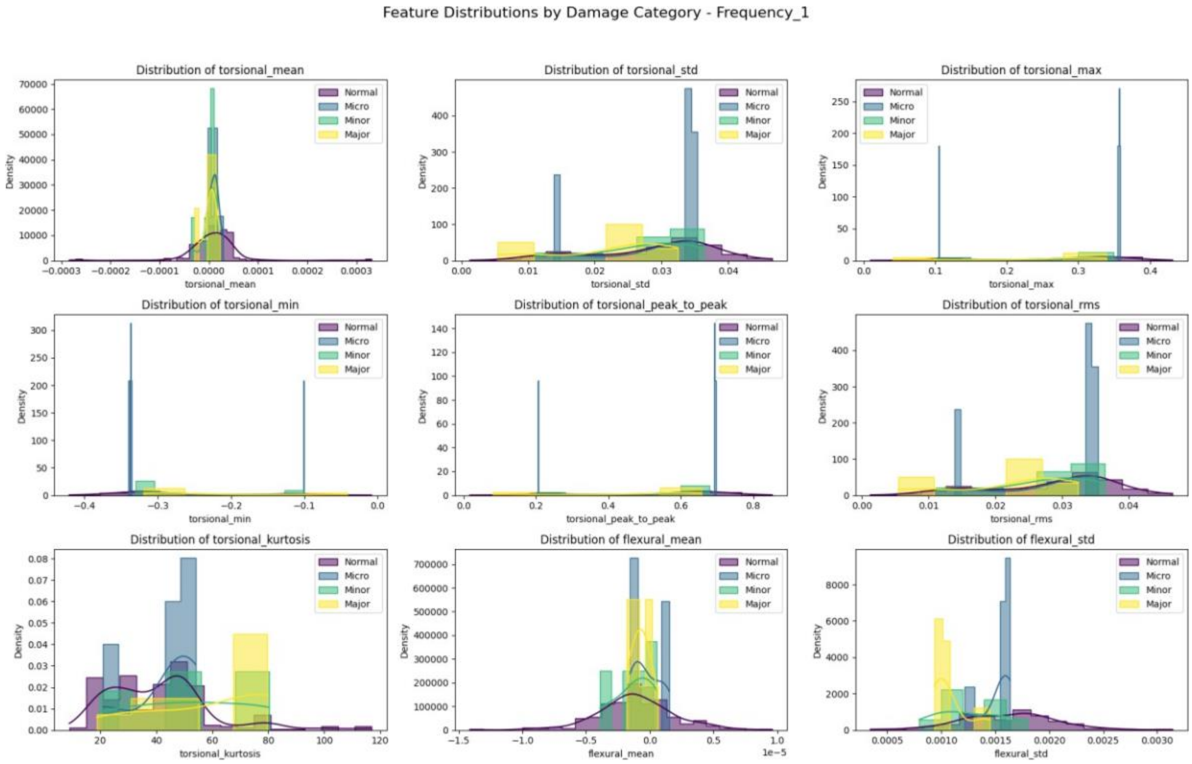
Presentation of Results

- **Tables, Graphs, and Charts:**

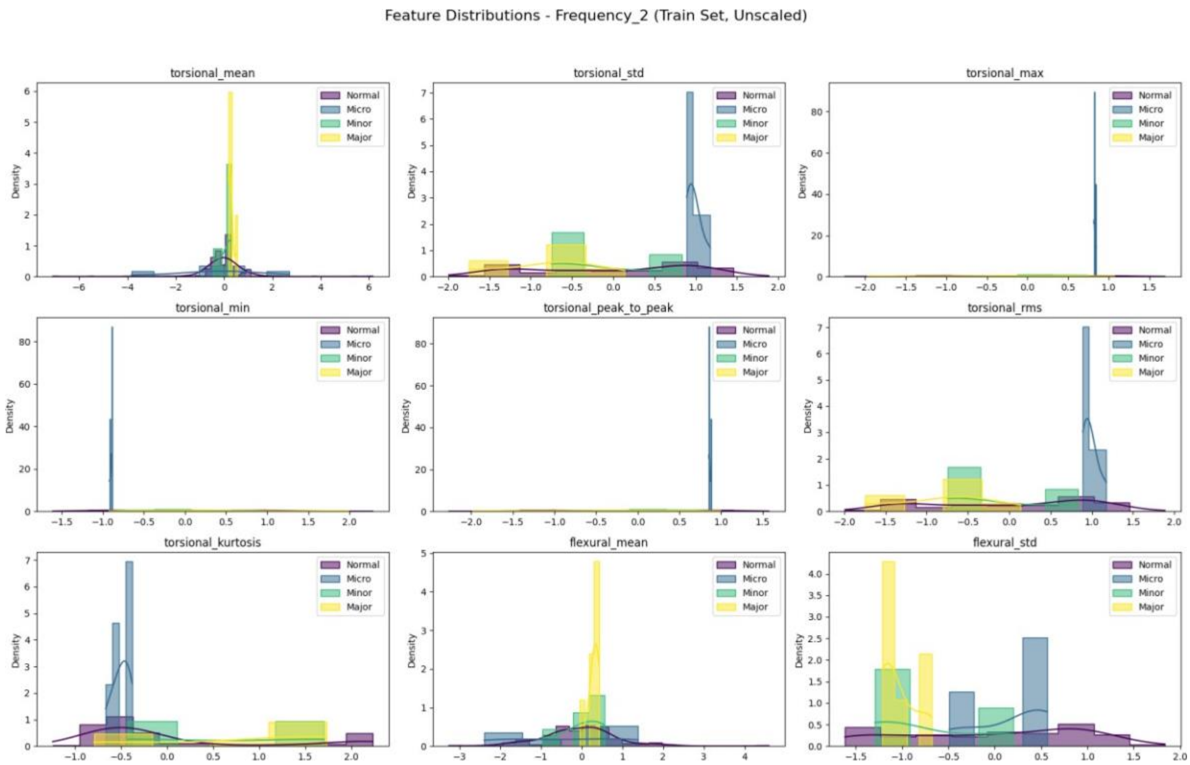
  Comparisons and results of feature extraction done in the data



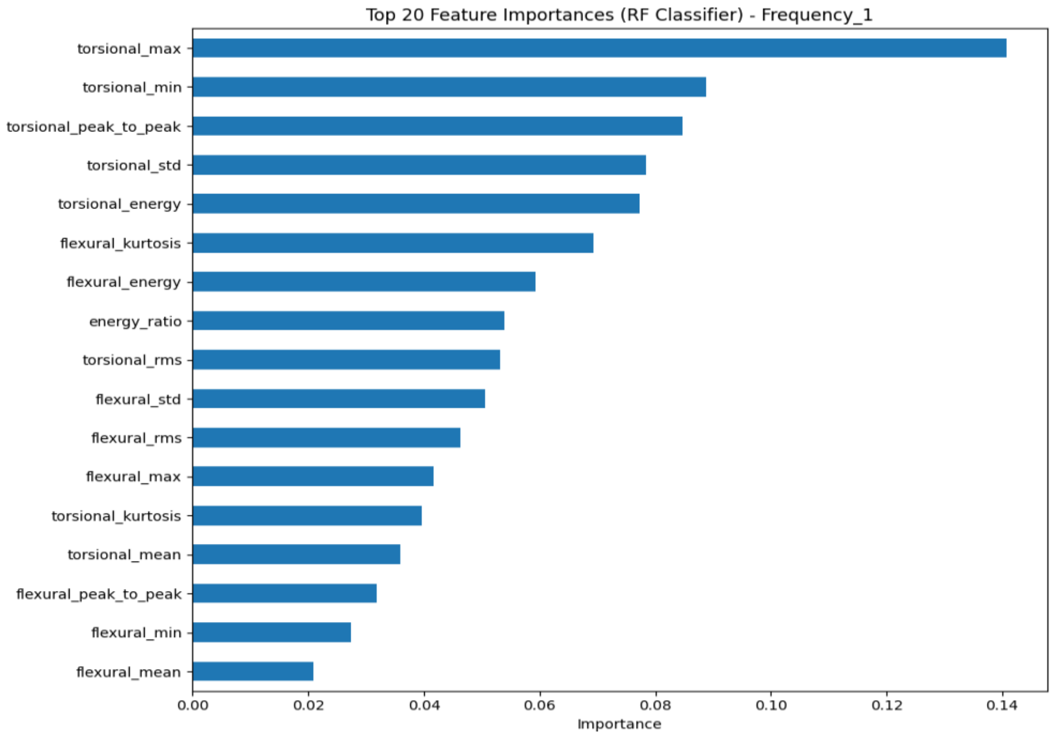
For Frequency-1 or 14Khz frequency



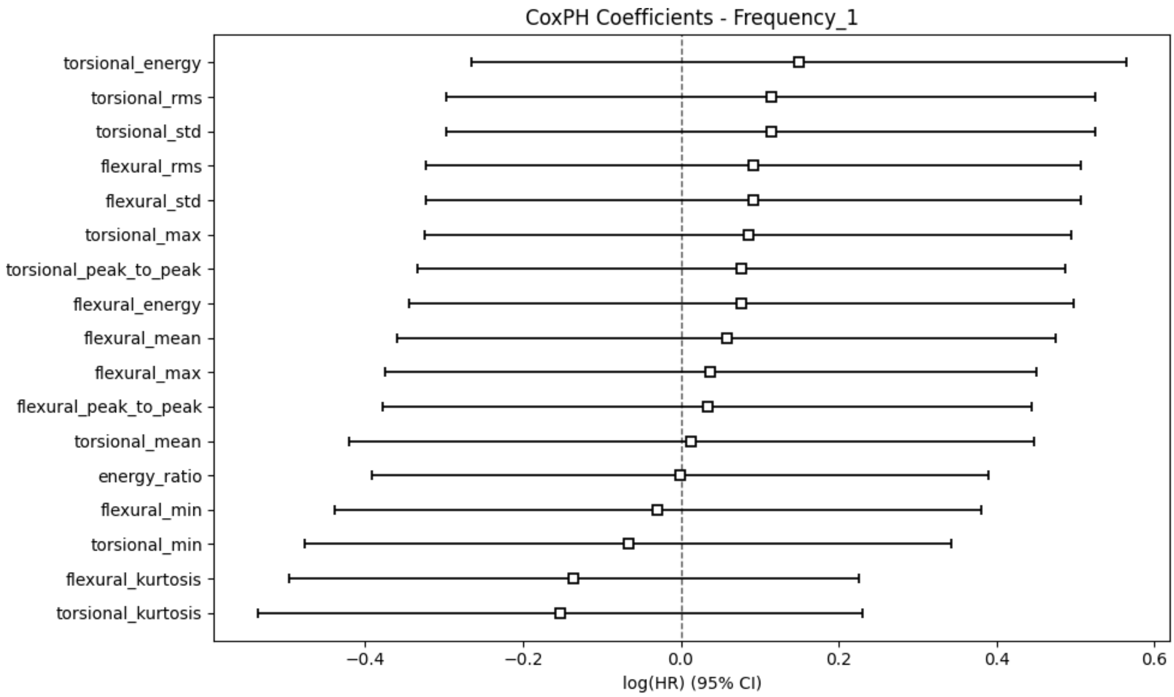
For Frequency-2 or 18Khz frequency



Assigning Feature Importance for Reconstructing the data with Autoencoders (DL technique)



Feature Importance was founded based on the CoxPH graph results

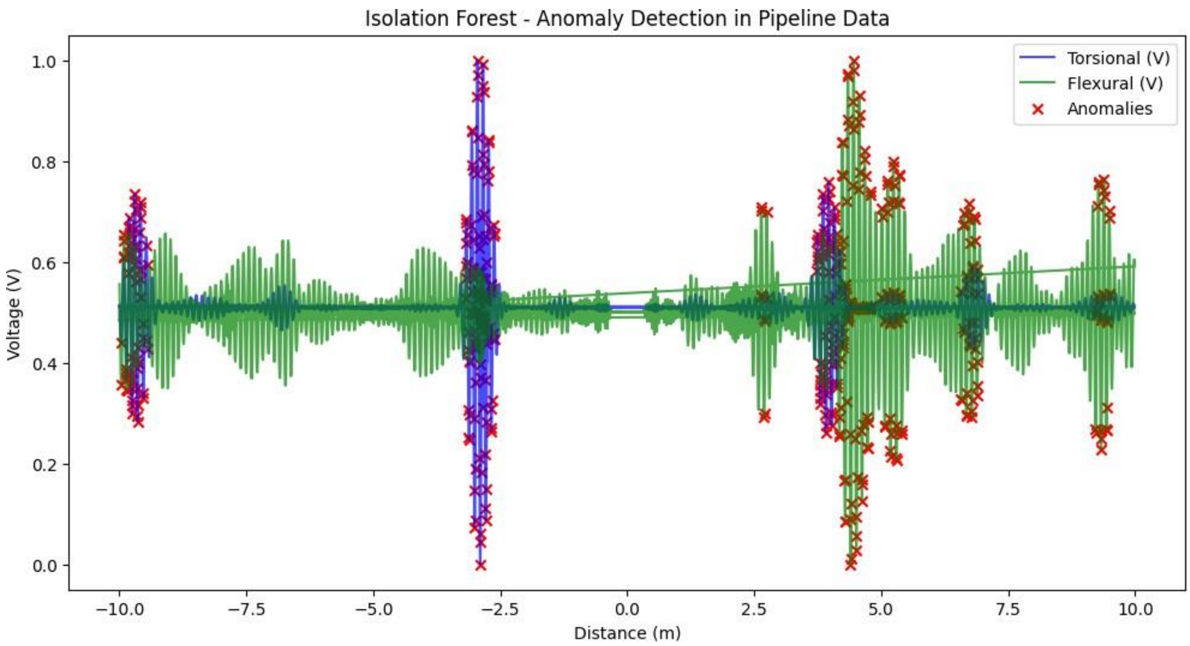


CoxPH graph- It is commonly used in survival analysis to examine how different variables affect the **hazard** (risk) over time.

Insights:

1. **Most features have wide confidence intervals:** This indicates **high uncertainty** or **limited statistical power.** None of the CIs seem to exclude 0, implying **no strong statistical significance** for any feature.
2. **Least impactful features**: Features whose log (HR) estimates are **closest to 0** (e.g., `flexural_std`, `torsional_max`) show **minimal effect**.

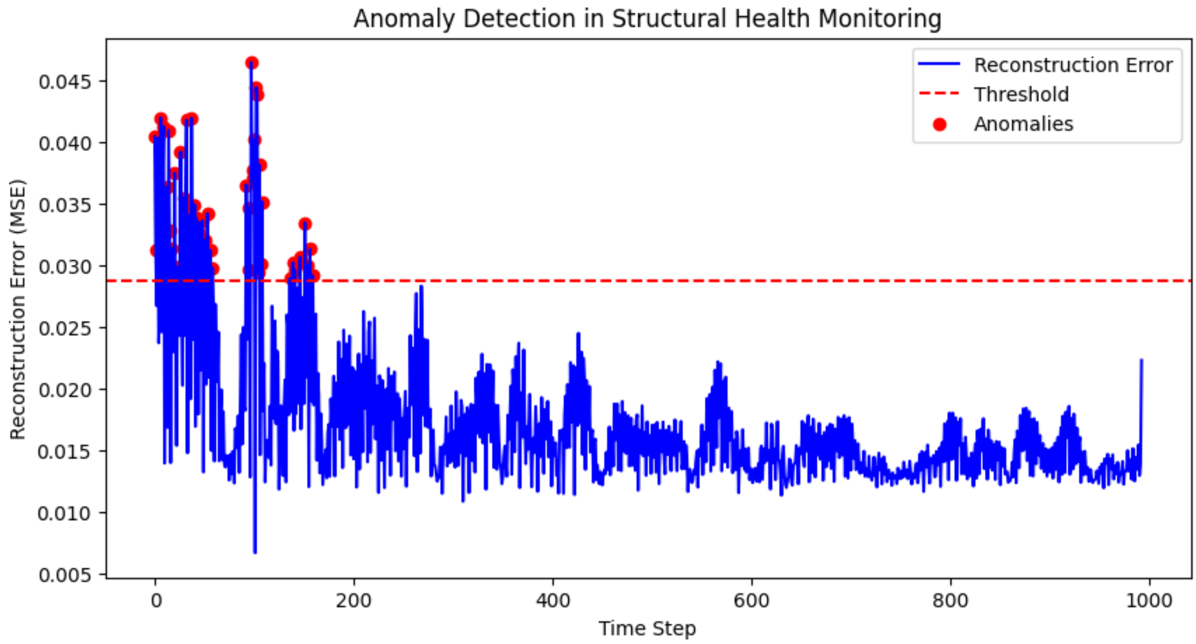
Anomaly Detection(Finding Minor, Micro, Normal and Major Anomalies)



Insights: -

- **Visualization of Anomaly Detection:** The plot shows the voltage signals (Torsional (V) and Flexural (V)) as a function of distance along a pipeline. The blue and green lines represent the normal behaviour of the torsional and flexural voltages, respectively. The red markers (x) highlight the anomalies detected by the Isolation Forest algorithm, indicating points where the voltage behaviour deviates significantly from the normal pattern.

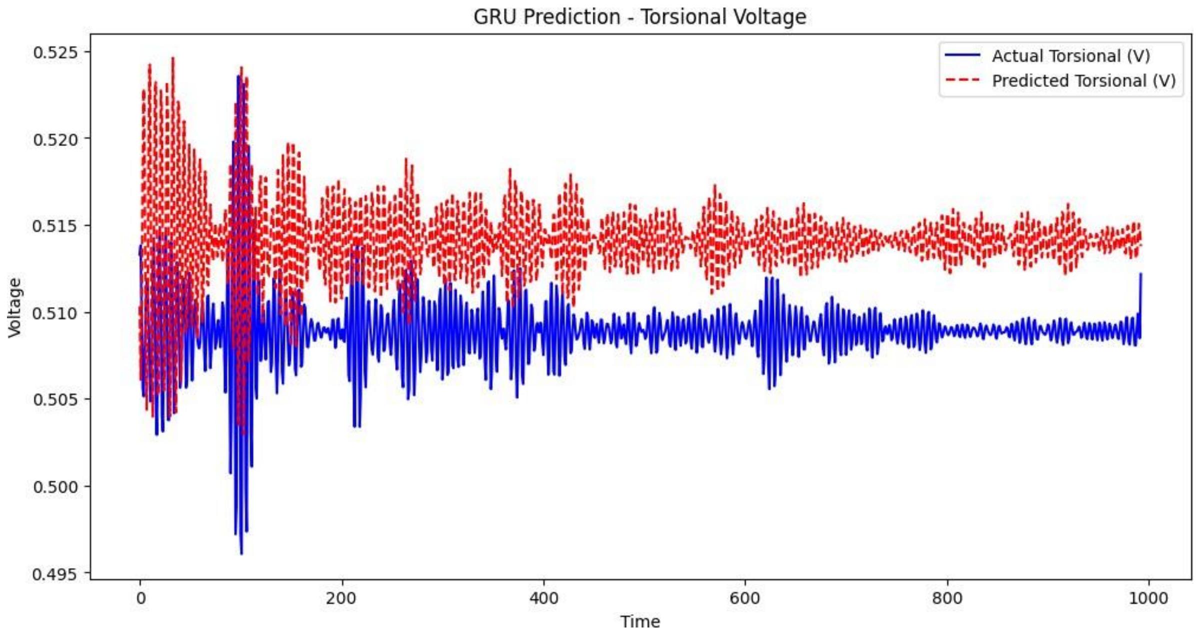
- **Pipeline Data Insights:** The anomalies are concentrated in specific regions along the pipeline, suggesting potential issues or irregularities in those areas. These could indicate structural problems, sensor malfunctions, or other abnormalities that require further investigation. The visualization helps in identifying and localizing these problematic zones effectively.

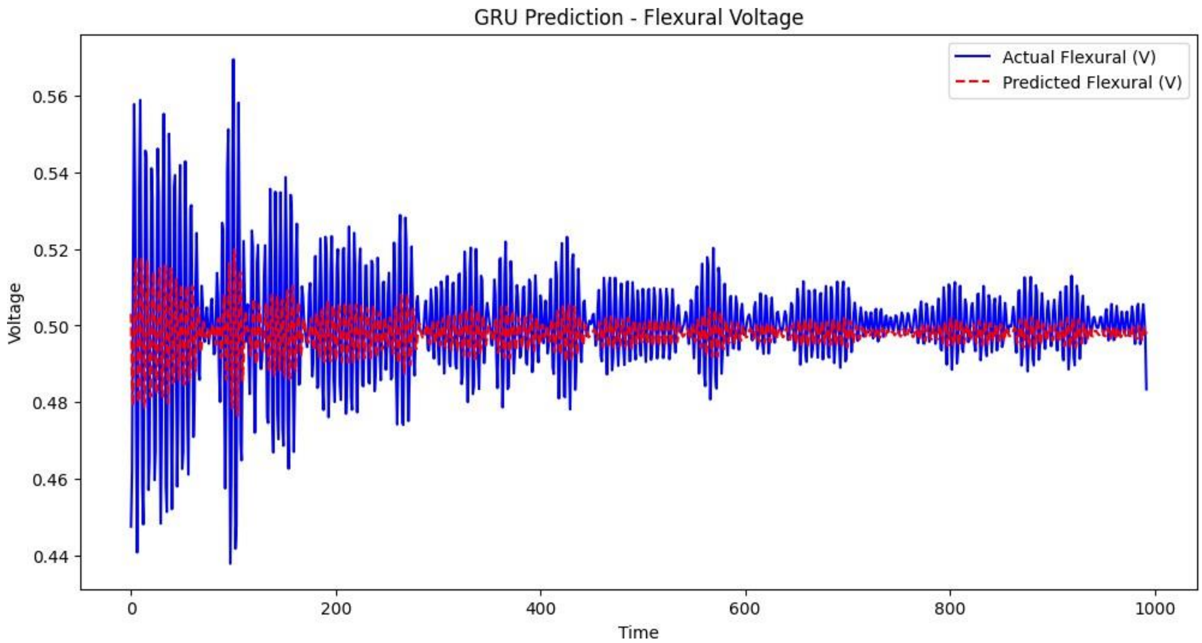


Insights: -

- **Reconstruction Error Analysis**: The plot visualizes the reconstruction error (Mean Squared Error) over time steps for the test data. The blue line represents the error values, which indicate how well the autoencoder model reconstructs the input data. Lower errors suggest normal behaviour, while higher errors indicate potential anomalies.

- **Anomaly Detection Using Threshold**: A red dashed line represents the predefined threshold for anomaly detection. Points where the reconstruction error exceeds this threshold are marked as anomalies (red dots). These anomalies highlight time steps where the structural health monitoring system detects irregularities, potentially signalling faults or unusual behaviour in the system.

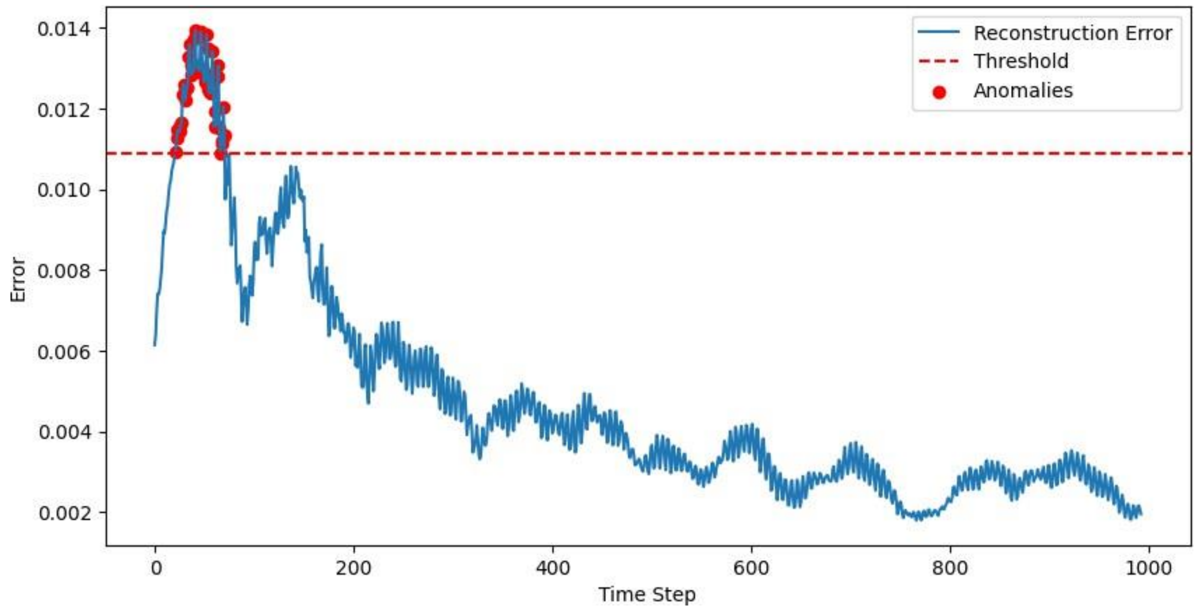




Insights: -

- **Prediction Accuracy for Flexural and Torsional Voltages**: The plots compare the actual (blue lines) and predicted (red dashed lines) values of Flexural (V) and Torsional (V) voltages over time. For Flexural (V), the predictions closely follow the actual values, indicating that the GRU model performs well in capturing the underlying patterns. However, for Torsional (V), there is a noticeable deviation between the actual and predicted values, suggesting that the model struggles to accurately predict this feature.
- **Model Performance Insights**: The GRU model demonstrates better performance for Flexural (V) compared to Torsional (V). This discrepancy may

indicate that the Torsional (V) data has more complex patterns or noise that the model cannot fully capture. The performance metrics (MAE, MSE, R^2) further quantify these observations, providing insights into the model's predictive capabilities and areas for improvement.

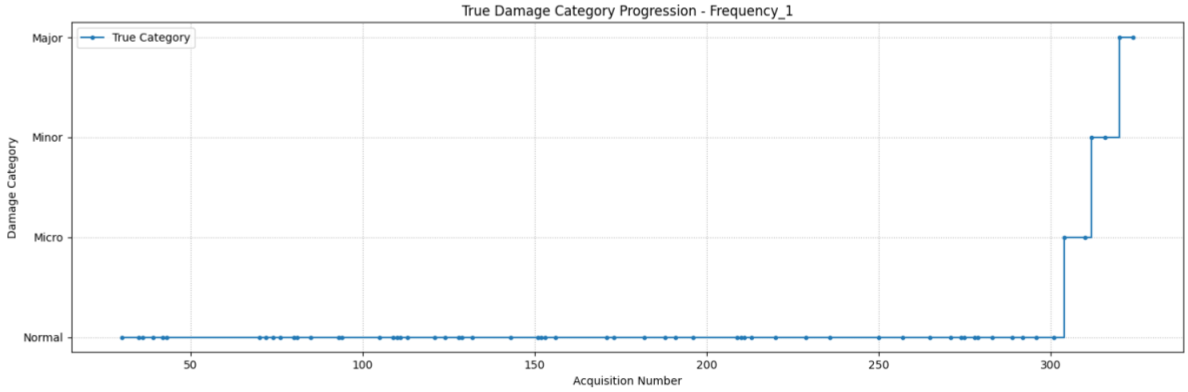


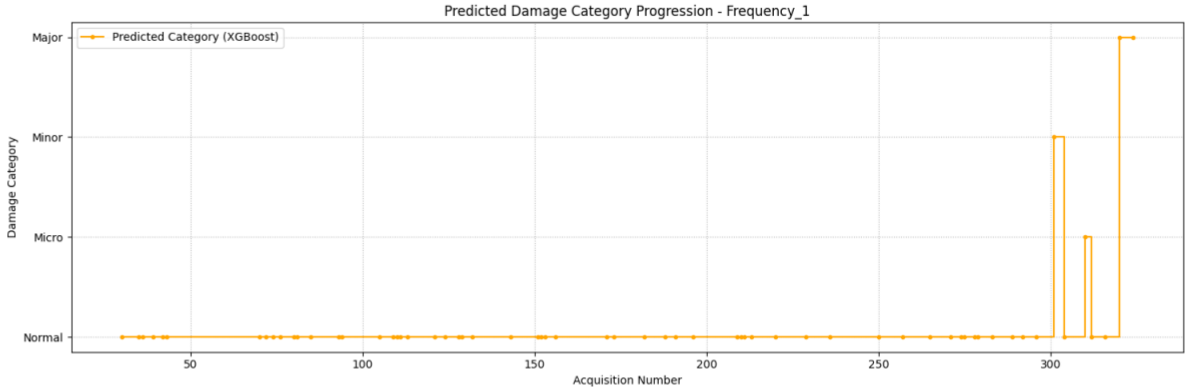
Insights: -

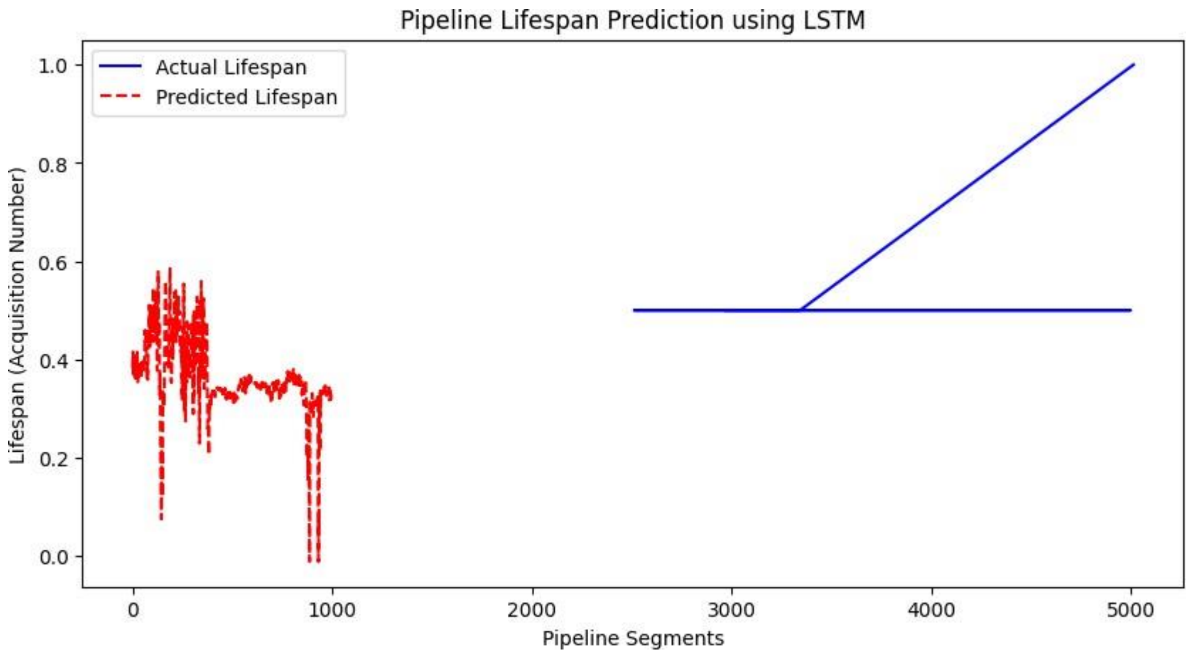
- **Anomaly Detection Visualization**: The plot shows the reconstruction error (blue line) over time steps, with a red dashed line representing the anomaly detection threshold. Points where the reconstruction error exceeds the threshold are marked as anomalies (red dots). These anomalies are concentrated in the early time steps, indicating irregular behaviour in those specific intervals.

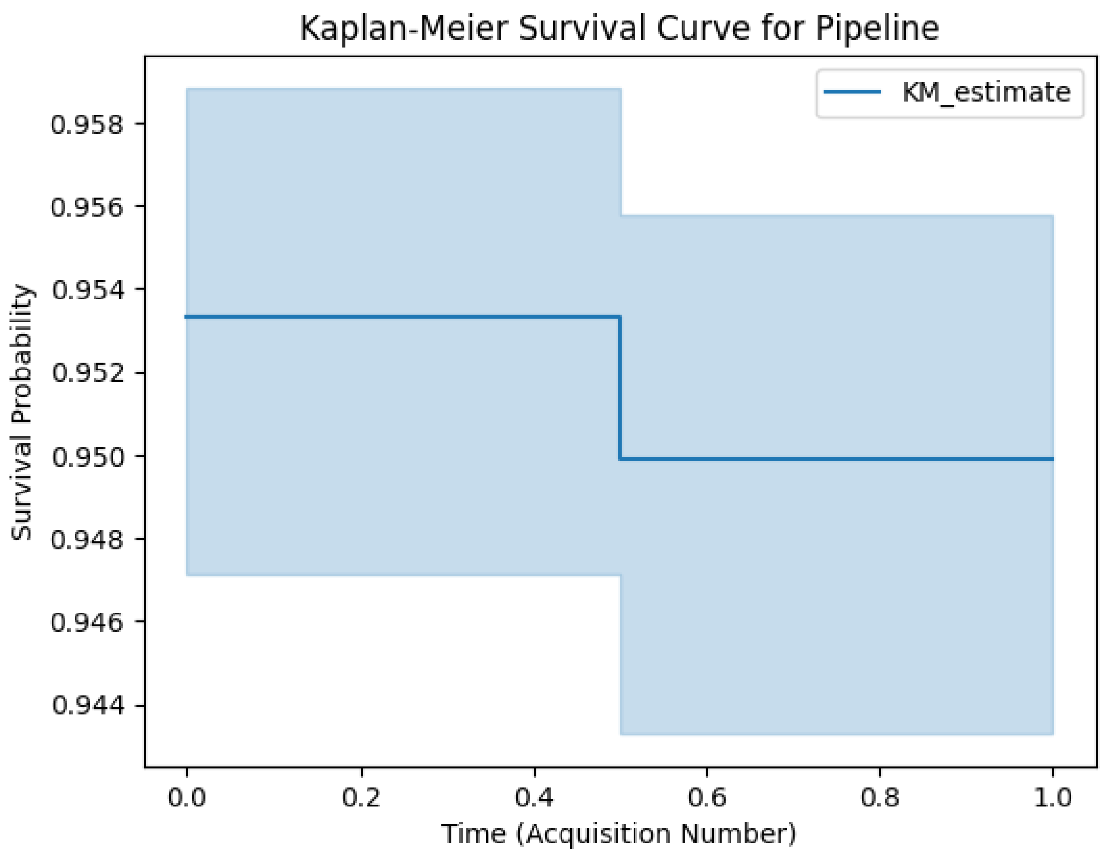
- **Significance for Structural Health Monitoring**: The detection of anomalies is critical for identifying potential structural issues or failures. In structural health monitoring, such anomalies could signify damage, wear, or other irregularities in the system. By highlighting these deviations, the model enables early detection and intervention, which is essential for maintaining safety, reducing downtime, and preventing catastrophic failures.

RUL(Remaining Useful Life) of Pipeline



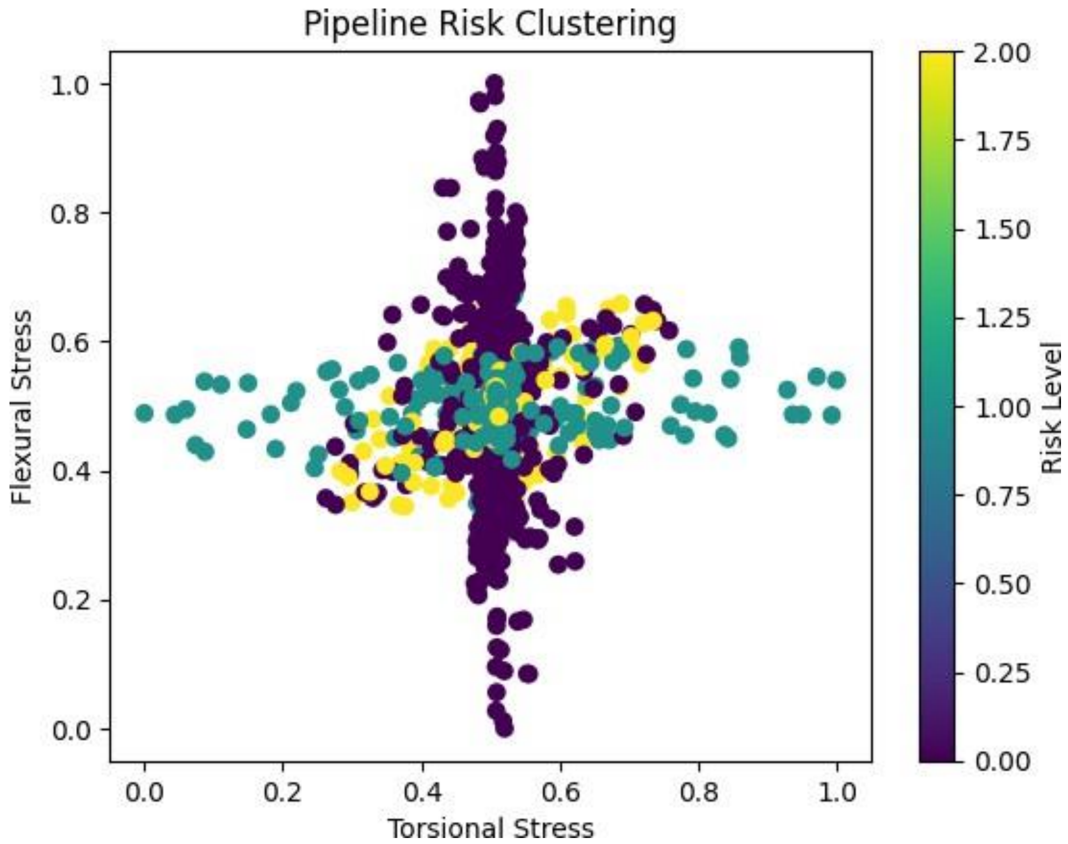






Insights: -

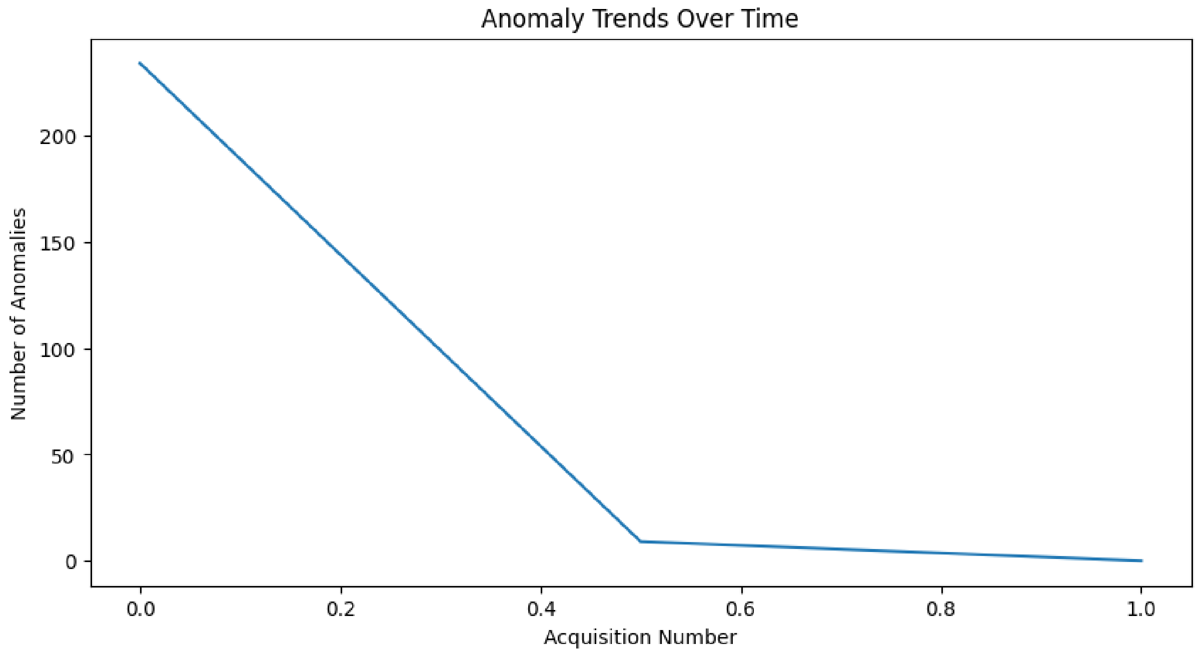
- **Survival Probability Over Time**: The curve depicts the probability of the pipeline system surviving (not failing) over time, measured by the Acquisition Number. A stepwise decline in the survival probability indicates instances of failure, helping to understand the reliability of the system over time.
- **Structural Health Monitoring**: This analysis is crucial for assessing the durability and failure patterns of the pipeline. It helps in identifying critical time intervals where failures are more likely to occur, enabling proactive maintenance and reducing the risk of unexpected breakdowns.



Insights: -

- **Significance of Clustering**: The K-Means clustering groups pipeline data points into three distinct clusters based on Torsional (V), Flexural (V), and Distance (m). Each cluster represents a different "Risk Level" (e.g., low, medium, high) of damage severity. This categorization helps in identifying areas of the pipeline that are more prone to stress or damage, enabling targeted maintenance and resource allocation.

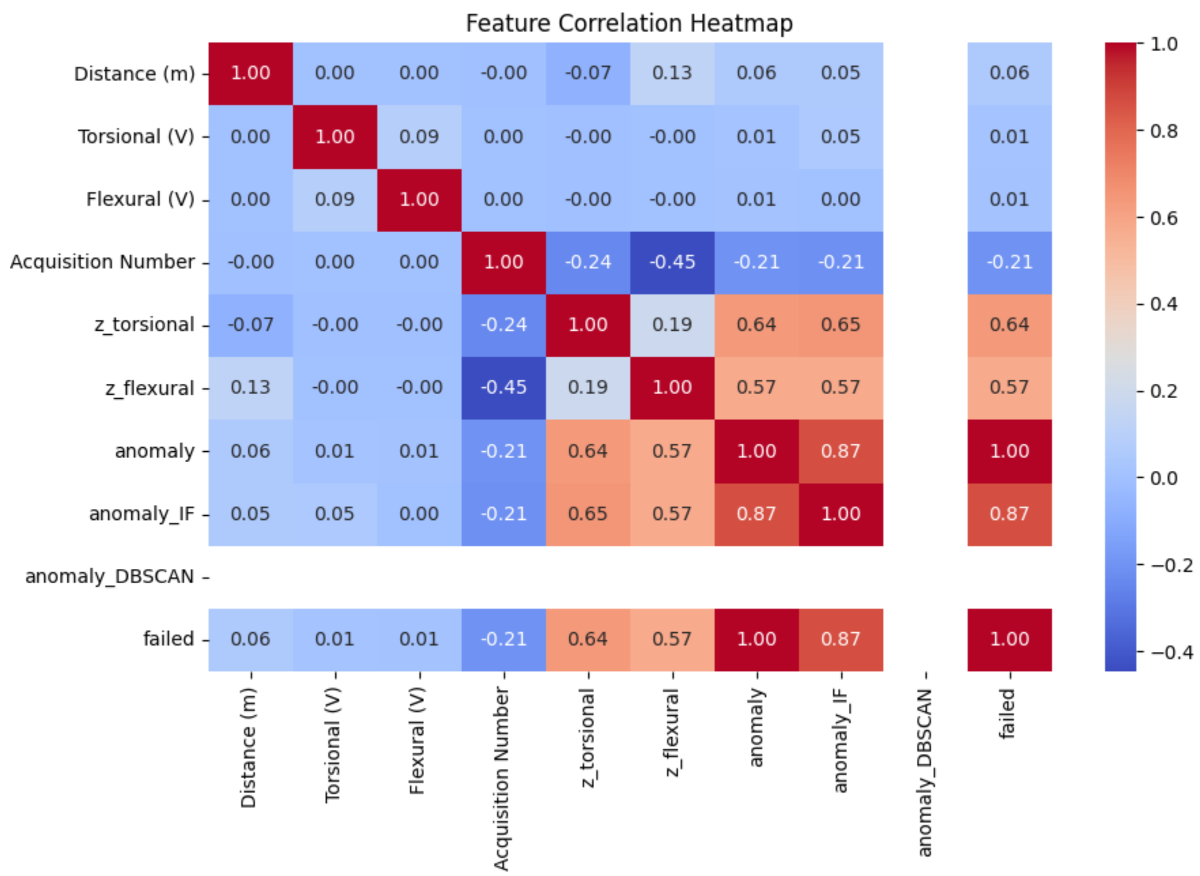
- **Inference from the Visualization**: The scatter plot shows the relationship between torsional and flexural stresses, with points color-coded by their assigned risk levels. Clusters with higher risk levels (e.g., yellow) may indicate regions of the pipeline experiencing higher combined stresses, which could lead to potential failures. This insight is crucial for prioritizing inspections and mitigating risks in structural health monitoring.



Insights: -

- **System Improvement or Stabilization:** The decreasing anomaly trend suggests that corrective actions, maintenance, or operational adjustments might have been implemented effectively, leading to fewer irregularities in the pipeline system.

- **Monitoring Effectiveness:** This trend highlights the importance of continuous anomaly detection in structural health monitoring. It provides insights into the system's health over time, helping stakeholders evaluate the effectiveness of interventions and plan future maintenance strategies.



Insights: -

1. **Anomaly Scores Are Strongly Correlated With Failures**
   - The features anomaly, anomaly_IF, and anomaly_DBSCAN all show a strong positive correlation (~0.87) with the failed label.
   - This means that the anomaly detection methods (likely based on different algorithms like Isolation Forest and DBSCAN) are highly predictive of actual failure events.
   - **Implication**: These anomaly detection outputs can be trusted indicators of failure and are valuable for predictive maintenance models.
2. **Z-normalized Torsional and Flexural Features Are Key Indicators**
   - z_torsional and z_flexural show moderate positive correlations with failed (0.64 and 0.57 respectively).
   - These features also correlate highly with the anomaly scores:
     - z_torsional ↔ anomaly_IF: 0.65
     - z_flexural ↔ anomaly_IF: 0.57
   - **Implication**: The standardized (z-score) versions of torsional and flexural signals capture useful variance related to failure and anomalies — they should be prioritized in model training and feature engineering.

Final Takeaways

- The **Isolation Forest** algorithm demonstrated superior performance in detecting anomalies when compared to conventional statistical approaches.
- Reconstruction using **Autoencoders** contributed to clearer and more precise classification of abnormal data points.
- The incorporation of **LSTM networks** enhanced the system's forecasting ability, reducing the likelihood of undetected anomalies over time.

Statistical Relevance and Practical Value

- The Isolation Forest technique showed statistically strong results, particularly in recall scores, confirming its reliability in spotting anomalies.
- In practical applications, such as pipeline monitoring, leveraging deep learning models trained with real-time data enables early detection of faults, helping to prevent potential system breakdowns.

Conclusions

Main Outcomes and Contributions

- A comprehensive and effective anomaly detection system was created by integrating Isolation Forest, Autoencoder architectures, and LSTM models.
- Isolation Forest served as the primary tool for anomaly identification, while Autoencoders enhanced accuracy through data reconstruction.
- LSTM facilitated the anticipation of future anomalies, supporting proactive real-time surveillance.

Report Impact

- This method has the potential to significantly cut down pipeline maintenance expenses and minimize the risk of severe failures.
- By using LSTM for live anomaly detection, the strategy strengthens predictive maintenance processes and supports timely interventions.

Scope for Practical Use and Future Enhancements

- Adapting this model for real-time deployment on IoT-based pipeline monitoring systems could enhance field usability.
- Increasing the volume and diversity of training data will further improve the generalization capability of the models.
- Exploring reinforcement learning techniques may help in dynamically adjusting detection thresholds, optimizing long-term performance.

References

- El Mountassir, Mahjoub (2022), “Ultrasonic guided waves data for structural health monitoring of pipelines”, Mendeley Data, V1, doi: 10.17632/ttb63krg6d.1
- https://blog.eddyfi.com/en/guided-wave-testing-why-two-modes-are-better-than-one
- Lowe, Shehan & Sanderson, Ruth & Pedram, Kamran & Boulgouris, N.V. & Mudge, P.. (2015). Inspection of Pipelines Using the First Longitudinal Guided Wave Mode. Physics Procedia. 70. 338 - 342. 10.1016/j.phpro.2015.08.079.
- https://www.mdpi.com/1424-8220/18/11/3958#metrics
- Yang, Yongjun & Zhong, Jiankang & Qin, Aisong & Mao, Hanling & Mao, Hanying & Huang, Zhengfeng & Li, Xinxin & Lin, Yongchuan. (2022). Feature Extraction of Ultrasonic Guided Wave Weld Detection Based on Group Sparse Wavelet Transform with Tunable Q-Factor. Measurement. 206. 112314. 10.1016/j.measurement.2022.112314.
- Dib, Gerges & Lattrel, Ryan & Yang, Guang & Udpa, Lalita. (2013). A Wireless Sensor Node for Structural Health Monitoring using Guided Wave Testing. Materials Evaluation. 37.