A Comparative Study of Machine Learning Techniques for AE-Based Corrosion Detection with Emphasis on Transformer Models

Osama Shahid Ali, Lukman B A Rahim
Center for Cyber Physical Systems, Universiti Teknologi PETRONAS, Seri Iskandar, Malaysia

***Abstract*—Corrosion-induced damage poses a critical threat to the structural integrity of fluid transport pipelines, necessitating advanced detection strategies for early intervention. This study investigates the use of acoustic emission (AE) monitoring in conjunction with machine learning techniques to identify anomalies indicative of corrosion. A comprehensive analysis of supervised, unsupervised, semi-supervised, and self-supervised learning methods is presented, with emphasis on their suitability for AE-based anomaly detection. Building upon this foundation, we implement and evaluate multiple machine learning models—including K-Nearest Neighbours (KNN), Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNN)—and compare them to a Transformer-based model integrated into a hybrid CNN-Transformer architecture. Experimental results demonstrate that the hybrid model outperforms all baselines, achieving R-squared values of 0.7037 for Acoustic Signal Level (ASL) and 0.6836 for Root Mean Square (RMS), thus confirming its superior ability to capture both local and long-range dependencies in acoustic emission data. A systematic review of recent Transformer-based corrosion detection models further contextualizes the results. This research highlights the promise of Transformer-based models in robust, real-time corrosion monitoring and offers a pathway toward more intelligent, machine learning-driven infrastructure maintenance systems.**

***Keywords*—Acoustic emissions; transformer based models; machine learning**

I. Introduction

The structural integrity of pipelines is critical to the safe and efficient transport of fluids across industrial infrastructure. Among the various threats to pipeline reliability, corrosion remains one of the most pervasive and damaging phenomena, often leading to catastrophic failures, economic losses, and environmental hazards. As such, early and accurate detection of corrosion-related anomalies is essential for preventive maintenance and risk mitigation. Traditional inspection methods, such as ultrasonic testing, visual inspection, or radiography, are often constrained by accessibility, resolution, and cost. These limitations have catalysed the shift toward non-destructive monitoring approaches, particularly those based on acoustic emission (AE) signals, which can detect internal damage mechanisms in real-time without interrupting operations.

Within this context, the field of anomaly detection has gained considerable traction. By identifying deviations from normal behaviour in sensor data, anomaly detection methods enable the recognition of subtle, early-stage corrosion signatures. These methods span from classical statistical techniques to advanced machine learning and deep learning models. More recently, the adoption of Transformer-based architectures—originally developed for natural language processing—has opened new avenues for capturing long-range dependencies and intricate temporal patterns in time-series data, including AE signals. Transformers, especially when combined with Convolutional Neural Networks (CNNs), have demonstrated superior performance in learning both localized and global representations, making them particularly well-suited for signal-based corrosion detection.

This study contributes to this evolving landscape by applying and evaluating a hybrid CNN-Transformer model for AE-based corrosion anomaly detection in pipelines. Unlike conventional approaches that often struggle with noisy, high-dimensional, and unlabelled signal data, the hybrid model leverages CNNs for localized feature extraction and Transformer encoders for contextual temporal modelling. A comprehensive comparison against classical machine learning algorithms and other deep learning models reveals the hybrid model's significant performance advantage in predicting key corrosion indicators such as Acoustic Signal Level (ASL) and Root Mean Square (RMS).

Furthermore, this work contextualizes its findings within the broader literature by examining prior studies on Transformer-based corrosion detection across various domains, including bridge inspection, UAV imagery, pipeline surveillance, and 3D point cloud segmentation. By bridging insights from both image-based and signal-based corrosion detection, this paper not only validates the effectiveness of Transformer models in AE applications but also lays the groundwork for future self-supervised learning strategies in industrial health monitoring systems.

II. Background of the Study

A. Anomaly Detection

Anomaly detection is the process of identifying data points, patterns, or events that significantly deviate from the normal or expected behaviour within a dataset [1]. These deviations, known as anomalies or outliers, can indicate critical incidents such as errors, defects, fraud, or other unusual activities. Anomaly detection is used across various fields like cybersecurity, finance, manufacturing, and healthcare to detect problems early and prevent larger issues. It typically involves building a model of normal behaviour and then comparing new data to this model to find anything that stands out as abnormal.

Anomaly detection offers several significant advantages, especially in systems where early identification of unusual behaviour can prevent serious consequences. One of the primary benefits is its ability to uncover rare or previously unknown issues, such as hidden faults in machinery, undetected cyberattacks, or fraudulent financial transactions, which might not be identifiable using standard monitoring techniques [2]. It also supports automation in monitoring processes, reducing the need for manual oversight and allowing organizations to respond more quickly and efficiently to emerging problems. Furthermore, anomaly detection improves decision-making by providing insights into unexpected behaviour and enabling predictive maintenance, risk management, and enhanced security. Its flexibility across domains and ability to handle large volumes of data in real time make it a powerful tool for increasing operational reliability and safety.

B. Anomaly Detection Use Cases

Anomaly detection plays a pivotal role in cybersecurity, especially for identifying novel or stealthy attacks that signature-based systems often miss. For instance, a study on smart-city IoT networks applied federated and split learning methods to detect anomalies on the UNSW-NB15 dataset, achieving accuracy rates above 97% and reducing false positives in real-time intrusion detection. Another work introduced a PCC-CNN deep-learning model capable of detecting multi-class attacks in IoT environments using datasets like NSL-KDD and CICIDS2017, demonstrating robust performance in real-time classification [3], [4], [5]. These approaches exemplify how anomaly detection enhances threat detection in dynamic, decentralized network ecosystems.

In manufacturing, anomaly detection enables predictive maintenance by identifying early signs of equipment degradation. A machine-learning system using Random Forests and SVM, trained on sensor data like vibration, temperature, and pressure, successfully flagged anomalies indicative of impending motor and pump failures. This led to a remarkable 30% reduction in unplanned downtime and substantial cost savings. Additionally, a semi-supervised approach for pharmaceutical manufacturing equipment demonstrated its effectiveness in detecting failures in complex multivariate time series without requiring labelled fault data [6], [7]. These techniques shift maintenance strategies from reactive to proactive, improving operational efficiency and equipment lifespan.

Deep learning–based anomaly detection methods are transforming system reliability across industrial settings. One paper [8] explored autoencoders, RNNs, and CNNs to examine sensor streams in a manufacturing environment. The proposed deep-learning framework outperformed traditional techniques, enabling earlier detection of equipment faults. Another study [9] implemented a federated learning setup using 1D-CNN and bi-LSTM to process time-series sensor data across distributed devices, achieving 97.2% test accuracy and enabling edge-based anomaly detection that preserves data privacy. Together, these works illustrate how advanced neural architectures facilitate real-time fault prediction and distributed diagnostics in Industry 4.0 systems.

In civil and structural engineering, anomaly detection methods are increasingly used to identify corrosion on critical infrastructure—like bridges, towers, and building façades—using aerial imagery [10] collected by drones or unmanned aerial vehicles (UAVs). For instance, CorrDetector [11] is an ensemble deep learning model based on convolutional neural networks (CNNs) that analyses drone-captured images of structures (e.g., telecom towers) to locate and identify corroded regions [12]. This framework achieved significantly better classification accuracy compared to prior state-of-the-art models, enabling automated, remote, and scalable corrosion monitoring in hard-to-reach areas. Such anomaly-driven detection enhances safety, reduces inspection costs, and enables timely maintenance decisions [13].

C. Types of Anomalies

Point anomalies refer to individual data instances that deviate significantly from the majority of the data and can be identified as unusual regardless of the context in which they occur [14]. These anomalies are the simplest and most intuitive form, appearing as sharp spikes or dips in data distributions. A point anomaly stands alone as being inconsistent with the expected pattern—there is no need to examine its surrounding values or additional variables [15]. For example, in banking, if a customer typically spends less than $100 per transaction, but one day a transaction of $10,000 is recorded, that transaction is likely to be a point anomaly. Similarly, in a medical dataset, a sudden and extreme heart rate recorded in an otherwise stable trend could be flagged as anomalous. These anomalies are typically rare but can be highly impactful, often signalling errors, fraud, or critical failures. Because they occur in a variety of domains—from finance and healthcare to environmental monitoring and manufacturing—point anomalies are often the first type considered when building anomaly-aware systems. Despite their simplicity, understanding and classifying them accurately is crucial due to the potential severity of the events they represent.

Contextual anomalies occur when a data point is considered normal in one context but anomalous in another. Unlike point anomalies, these require additional contextual information—such as time, location, or season—to assess whether the data instance is truly unusual [16]. A contextual anomaly cannot be identified solely by looking at the value itself; its significance emerges only when the surrounding environment or condition is considered [17]. For example, a temperature of 25°C might be perfectly normal during the summer but highly unusual during the winter, making it a contextual anomaly in the latter case. Another example would be a sudden drop in stock prices that is abnormal only when viewed against the typical patterns of that specific market on a certain day of the week. These types of anomalies are especially prevalent in time-series and spatiotemporal data, where the same value can be either expected or unexpected depending on when and where it occurs. Recognizing contextual anomalies is essential in fields such as climate science, behavioural analytics, and industrial monitoring, where environmental variability plays a key role in defining what is normal.

Collective anomalies refer to a set or sequence of data instances that are anomalous when considered together, even though the individual data points within the group may appear normal on their own. The anomaly lies not in the data points themselves, but in the pattern or relationship between them. This type of anomaly typically appears in sequential or structured data, such as time-series or spatial data grids [18]. For example, in network traffic analysis, a single packet may appear normal, but a burst of similar packets over a short duration might indicate a denial-of-service attack—a collective anomaly. Similarly, in health monitoring systems, a slight elevation in blood pressure over one reading might not be a concern, but a sustained increase over several hours could signal the onset of a medical emergency. These anomalies are crucial in identifying coordinated or evolving patterns that could go unnoticed when data is analysed in isolation. They are especially important in domains like cybersecurity, fault detection, and system diagnostics, where anomalies emerge not from individual values, but from the way groups of values interact over time or space.

D. Anomaly Detection Techniques

1) Statistical techniques are among the earliest and most fundamental approaches to anomaly detection [19]. These methods rely on the assumption that normal data points follow a known statistical distribution (e.g., Gaussian, Poisson), and anomalies are observations that deviate significantly from this distribution. Common statistical measures used include mean, variance, standard deviation, and probability density functions.

A classic example is the Z-score method, where any data point that lies beyond a threshold (commonly 3 standard deviations from the mean) is considered anomalous[20]. This approach works well for data that is symmetrically distributed, such as in many industrial process control settings. Grubbs' Test is another statistical test used for identifying outliers in a univariate dataset that assumes normality; it tests the hypothesis that the extreme value in the dataset is an outlier [21].

Another foundational statistical method is the Interquartile Range (IQR) approach, which defines outliers as data points lying below $Q1-1.5\times IQR$ or above $Q3+1.5\times IQR$, where Q1 and Q3 are the first and third quartiles, respectively [19]. This is particularly useful for skewed distributions and boxplot-based anomaly visualization.

More advanced techniques include Kernel Density Estimation (KDE), a non-parametric way to estimate the probability density function of a random variable [22]. Anomalies are identified as points in regions of low estimated density. KDE is flexible and does not require assumptions about the distribution form.

In time-series applications, statistical methods like Moving Average and Exponential Smoothing are employed to detect deviations from expected trends. These are often combined with control limits to flag values that exceed expected variation. Similarly, Cumulative Sum Control Charts (CUSUM) and Exponentially Weighted Moving Average (EWMA) charts are widely used in quality control to detect small shifts in process mean [20].

These techniques are particularly valuable due to their interpretability, low computational cost, and ease of implementation. They are commonly used in industries such as finance (to detect fraud), manufacturing (to monitor production quality), and environmental science (to identify climate anomalies).

2) Time-series anomaly detection involves techniques that account for the temporal structure and order of the data [23]. Unlike traditional static datasets, time-series data includes observations collected over time, making the detection of anomalies more challenging due to seasonality, trends, and autocorrelation.

One of the most traditional techniques is the Autoregressive Integrated Moving Average (ARIMA) model. It predicts future values based on past observations. Anomalies are detected by comparing actual values to ARIMA-based forecasts; significant deviations signal potential anomalies. ARIMA is particularly effective when the data shows clear trends or periodic patterns.

Seasonal Decomposition of Time Series (STL) separates a time series into seasonal, trend, and residual components. Anomalies are typically identified in the residual component. This method helps isolate irregularities that cannot be explained by seasonality or trend.

Change-point detection techniques aim to identify points in time where the statistical properties of the series change [24]. These can include abrupt shifts in mean, variance, or autocorrelation structure. Algorithms like Bayesian Change Point Detection, Pruned Exact Linear Time (PELT), and Binary Segmentation are used in applications such as financial market analysis and equipment failure prediction [25].

More recent developments include the Matrix Profile method, which enables fast, scalable similarity search and anomaly detection in large time-series datasets [26]. By identifying the least similar sub sequences, Matrix Profile effectively highlights unusual patterns without requiring a model to be trained.

Time-series specific techniques are essential in fields like energy consumption forecasting, stock market analysis, industrial equipment monitoring, and healthcare analytics, where anomalies often follow complex temporal dynamics.

3) Signal-based anomaly detection techniques are widely used in domains where data is collected in the form of continuous waveforms, vibrations, or frequency-based signals. These techniques focus on identifying irregularities in the shape, amplitude, or frequency components of the signal data, which may indicate faults, noise, or other abnormal conditions. Applications include structural health monitoring, acoustic emission testing, vibration analysis in machinery, and biomedical signal processing such as ECG and EEG monitoring.

One common method is spectral analysis, where the frequency content of a signal is examined using tools like the Fourier Transform [27]. Anomalies may manifest as sudden spikes or drops in power at certain frequencies. In rotating machinery, for instance, deviations from baseline frequency signatures often suggest wear, imbalance, or bearing failure.

Another key method is wavelet transform analysis, which provides both frequency and time localization of signal features [28]. This is particularly useful for identifying transient anomalies, such as sudden impacts or short-duration faults. Wavelets can decompose signals at multiple scales, making them suitable for multi-resolution analysis.

Envelope analysis is another powerful technique used to detect impacts and modulated signals hidden within a complex waveform [29]. It's often used in gear and bearing diagnostics, especially where periodic impulses are masked by noise.

Cross-correlation and coherence techniques are used to compare signals across multiple channels or sensors to detect spatially coherent anomalies, such as propagating cracks in structures.

In AE-based corrosion monitoring, signal-based techniques can distinguish between noise, crack growth, and corrosion signals based on parameters like rise time, amplitude, duration, and energy [30]. These methods are critical in environments where traditional techniques may not detect internal flaws until catastrophic failure.

Signal-based methods are valued for their sensitivity and ability to capture minute signal changes that precede larger failures. However, they require high-fidelity data acquisition systems and advanced filtering to isolate meaningful features.

4) Machine learning techniques form a core component of modern anomaly detection systems. These techniques leverage data-driven learning models to discover complex and nonlinear relationships between features that can indicate abnormal patterns. Depending on the availability of labelled data and the structure of the problem, machine learning techniques are broadly categorized into four main types: supervised, unsupervised, semi-supervised, and self-supervised learning. Each of these categories provides different advantages and trade-offs, and their usage is dictated by the nature of the data and the application domain.

E. Machine Learning Anomaly Detection Techniques

1) Supervised learning involves training models on datasets that include labelled examples of both normal and anomalous data. These models learn to distinguish between the two classes based on the features present in the training data. Algorithms such as Support Vector Machines (SVM), Decision Trees, Random Forests, Logistic Regression, and Neural Networks are widely used in this category.

This approach can be highly effective when ample labelled data is available. It allows for the creation of highly discriminative models that can achieve strong generalization on unseen data [31]. However, its major limitation lies in the difficulty of acquiring sufficient labelled examples of anomalies, which are typically rare and diverse.

Supervised anomaly detection has been successfully applied in spam detection, medical diagnostics, and quality assurance systems where historical labelled datasets are available. Techniques such as cross-validation and synthetic data generation (e.g., SMOTE) are often used to handle class imbalance [32].

2) Unsupervised learning techniques are particularly useful when labelled data is unavailable, which is a common scenario in anomaly detection where anomalies are rare and difficult to label [1]. These models attempt to learn the inherent structure of the dataset without predefined labels, assuming that normal patterns dominate the data distribution. Anomalies are identified as instances that do not conform to these learned patterns.

Common approaches include clustering algorithms like k-means and DBSCAN, where outliers are data points that do not belong to any cluster or are far from cluster centroids. Dimensionality reduction techniques such as Principal Component Analysis (PCA) and Isolation Forest also fall under this category [14]. PCA reduces the data into a lower-dimensional space where anomalies can be observed as deviations from the principal components. One-Class SVM is another widely used method that attempts to learn the boundary of normal data and flags deviations as anomalies.

Unsupervised techniques are widely used in cybersecurity for detecting unauthorized access, in finance for fraud detection, and in manufacturing for identifying process faults. They require minimal prior knowledge about the data and are effective for exploratory analysis.

3) Semi-supervised learning sits between supervised and unsupervised methods. It utilizes a large amount of unlabelled data and a small amount of labelled data—usually only normal instances [2]. The model is trained to capture the distribution of normal data and detect deviations as potential anomalies.

A prominent approach in this category includes One-Class SVM, which learns a decision function that identifies the region in the feature space where the majority of the data lies. Autoencoders, a type of neural network, are also commonly used; they are trained to reconstruct normal data with high fidelity [33]. At inference time, high reconstruction errors suggest that the input does not conform to the learned normal pattern, signalling an anomaly.

Semi-supervised methods are well-suited to industrial applications, such as predictive maintenance, where anomaly data is limited. They are also used in biomedical monitoring, where only healthy data may be available for training.

4) Self-supervised learning (SSL) is an emerging paradigm in anomaly detection that leverages large volumes of unlabelled data by creating pseudo-labels through pretext tasks [34]. These pretext tasks are designed to train models to learn general data representations, which can then be used for downstream anomaly detection without requiring manually labelled anomalies.

In the context of anomaly detection, SSL often begins by defining surrogate tasks such as predicting masked input features (e.g., in tabular data), reconstructing corrupted signal segments (e.g., in time series), or predicting the temporal order of sequence segments [35]. Models trained on these tasks develop an internal representation of normal data distribution. During inference, deviations from expected patterns are flagged as anomalies.

In time-series anomaly detection, a common SSL strategy involves forecasting future values based on historical data using models like Transformers or Temporal Convolutional Networks. Large prediction errors are used as indicators of anomalous behaviour [36]. Similarly, contrastive learning—a subfield of SSL—maximizes agreement between different views or augmentations of the same data point while minimizing agreement with others, making it powerful for identifying rare deviations.

In industrial and biomedical applications, self-supervised pretraining has shown to improve anomaly detection performance [37]. For example, in acoustic emission-based corrosion monitoring, self-supervised pretraining allows the model to learn signal patterns unique to material degradation without needing explicitly labelled defect types.

SSL not only reduces the dependency on labelled data but also enhances robustness and generalizability across unseen data distributions. It has proven effective in fraud detection, predictive maintenance, healthcare monitoring, and cybersecurity, where anomalies evolve or emerge dynamically over time.

5) Deep learning techniques have revolutionized the field of anomaly detection by enabling systems to automatically extract hierarchical representations from raw, high-dimensional data. These techniques, primarily based on artificial neural networks, are capable of modelling complex and nonlinear data distributions, making them exceptionally powerful for detecting subtle and intricate anomalies.

Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are among the most commonly used deep learning models for anomaly detection. CNNs are particularly effective in applications involving image or spatial data, where anomalies may appear as localized irregularities. For instance, in industrial quality inspection, CNNs can identify manufacturing defects from high-resolution images [38]. RNNs, on the other hand, are designed to process sequential data and are extensively used in time-series anomaly detection, such as sensor data monitoring, by capturing temporal dependencies and flagging unexpected deviations.

Autoencoders are another widely used deep learning technique. These unsupervised neural networks are trained to reconstruct their input data. During testing, if the reconstruction error is high for certain inputs, those inputs are flagged as potential anomalies. Variants such as Variational Autoencoders (VAEs) and Denoising Autoencoders add probabilistic and noise-resilience components to improve detection robustness [39].

Generative Adversarial Networks (GANs) have also gained popularity for anomaly detection [40]. These models consist of a generator that creates fake data and a discriminator that tries to distinguish between real and generated data. The discriminator's inability to classify an instance properly can indicate that the data does not follow the expected distribution, thus signalling an anomaly. GAN-based methods have been applied in fields such as cybersecurity, medical imaging, and predictive maintenance.

Long Short-Term Memory (LSTM) networks, a type of RNN, are particularly effective for learning long-range dependencies in time-series data [41]. They have been employed in health monitoring systems, power grid analysis, and transportation networks to predict normal behaviour and detect anomalies through forecast errors.

Deep learning models benefit from large-scale datasets and computational resources, enabling them to outperform traditional methods in many scenarios [42]. However, they also present challenges such as model interpretability, high computational cost, and the need for careful hyperparameter tuning.

6) Transformers represent a significant advancement in machine learning models for anomaly detection due to their superior ability to model long-range dependencies and capture complex temporal patterns in data. Originally designed for natural language processing, Transformers have since been successfully adapted to various domains including vision, time-series analysis, and industrial monitoring [43]. Unlike traditional recurrent architectures like RNNs or LSTMs, Transformers rely on self-attention mechanisms that allow them to process all elements of a sequence simultaneously, leading to better scalability and improved performance, especially on large and noisy datasets.

One of the key advantages of Transformers in anomaly detection lies in their flexibility and robustness. They can be applied in both supervised and self-supervised settings and are capable of learning high-quality representations from unlabelled data. In comparison to CNNs and RNNs, Transformers offer superior performance in handling multivariate time-series data, where correlations across different features and timestamps must be considered jointly [44]. Moreover, self-attention helps in identifying subtle contextual anomalies that may span long sequences, which traditional models often miss.

This makes Transformers particularly valuable in industrial and infrastructure monitoring applications, such as in predictive maintenance and fault detection. One emerging application is anomaly detection in corrosion monitoring, especially using Acoustic Emission (AE) data, where annotated anomalies are difficult to obtain and the signals are often non-stationary and high-dimensional [45]. The ability of Transformers to learn from limited or unlabelled data and their capacity to model complex temporal dynamics makes them an ideal choice for such scenarios.

Furthermore, recent research has demonstrated that Transformer-based architectures can outperform conventional techniques in both accuracy and generalization when applied to corrosion-related anomaly detection tasks [46]. Their compatibility with self-supervised learning strategies enhances their effectiveness in low-resource settings, making them particularly relevant for pipeline health monitoring systems and structural integrity assessments.

The discussion of Transformer applications in corrosion detection, is expanded further in the next section, where we delve into their specific contributions, datasets, and models.

III. Prior Studies of Transformer Based Models for Corrosion Anomaly Detection

A. Comparative Analysis of Transformer Based Models for Corrosion Detection

The following paragraphs go into a deeper dive into each of the studies mentioned in Table I.

TABLE I. COMPARATIVE STUDY

Study	Transformer Model	Input Data	Learning Setup	Dataset	Accuracy/F1
Efficient Metal Corrosion Area Detection Model Combining Convolution and Transformer	Visual Transformer integrated in MCD-Net	Corrosion images (public dataset A and self-collected Dataset B)	Supervised Learning	Dataset A (AASHTO/BIRM corrosion levels), Dataset B (natural scene images)	F1 Score: 84.53%
CorFormer: a hybrid transformer-CNN architecture for corrosion segmentation on metallic surfaces	CorFormer (hybrid Transformer-CNN)	Corrosion images on metallic surfaces	Supervised Learning	Not stated (mentions 10 validation splits)	IoU improved by 2.7% over SOTA; FPS: 28
Civil Infrastructure Damage and Corrosion Detection: An Application of Machine Learning	Transformer block within CycleGAN (residual block-based Transformer)	UAV aerial images of bridges and steel infrastructure	Supervised Learning	1300 images (Bolte Bridge & Sky Rail, Victoria, AU)	F1 Score: 0.83; Accuracy: 0.989
A Multilevel Bridge Corrosion Detection Method by Transformer-Based Segmentation in a Stitched View	Transformer-based segmentation (e.g., SegFormer)	Stitched images of bridge surfaces	Supervised	Public dataset from Virginia Tech (440 images), augmented with Lundamo Bridge dataset (94 images)	F1 Score: 68.2% (improved from 60–61% with CNNs)
P-DETR: A Transformer-Based Algorithm for Pipeline Structure Detection	Pipe Detection Transformer (P-DETR, DETR-based with FNT module)	Aerial drone images of pipelines	Supervised	Custom dataset (2553 raw + augmented to 6127 images)	mAP: 55% (↑3 AP over DETR); Recall: 73.9%
Corrosion segmentation method of concrete drainage pipes based on point transformer	Point Transformer (U-Net-like)	3D point clouds from real + simulated concrete pipe data	Supervised Learning	200 real + 200 simulated 3D point cloud samples (Azure Kinect DK)	Accuracy: 94.36%; MIoU: 86.31%
SegFormer: Semantic Segmentation Based Transformers For Corrosion Detection	SegFormer	Bridge corrosion images from inspection reports (VDOT)	Supervised Learning	440 annotated images (AASHTO/BIRM guidelines)	Mean Acc: 81.39%; MIoU: 71.16%
Double-Attention YOLO: Vision Transformer Model for Transmission Line Fittings & Rust	Vision Transformer + Dual Attention YOLO	Hazy/defogged images of transmission line connection fittings	Supervised Learning	Not explicitly named; 7700 annotated images (Stage 2)	mAP@0.5:0.95 = 0.8674mAP@0.5 = 0.9948Macro-F1 = 0.661 (stage 1), 0.568 (stage 2)
Corrosion SAM: Adapting Segment Anything Model with Parameter-Efficient Fine-Tuning for Structural Corrosion Inspection	Segment Anything Model (SAM) + Vision Transformer (ViT-H) with Adapter modules (PEFT)	RGB corrosion images resized to 512×512; binarized	Supervised Learning	VDOT (440 images)KRDB (955 images)	VDOT: IoU 69.72%, Dice 81.04%KRDB: IoU 60.78%, Dice 72.97%
Pipeline defect recognition algorithm based on CNN-Transformer model for internal detection in PE gas pipelines	CNN + Transformer with Squeeze-and-Excitation (no positional encoding)	Internal pipeline defect images (1200), 1080×1080 px, collected via inspection robot	Supervised Learning	Custom dataset (1200 images) of PE pipeline defects	Accuracy: 97.8% (avg. over 10 runs)

The study titled "Efficient Metal Corrosion Area Detection Model Combining Convolution and Transformer" proposes MCD-Net, a hybrid deep learning architecture that integrates convolutional layers with a visual Transformer encoder to enhance corrosion detection on metallic surfaces [46]. By combining local feature extraction with global context modelling, the model effectively addresses challenges such as occlusions, lighting variability, and irregular corrosion patterns. It employs an attention-based multi-layer feature fusion mechanism and a multi-scale feature enhancement strategy to refine boundary segmentation. MCD-Net achieved an F1 score of 84.53% on a public dataset, demonstrating improved robustness to noise and lighting compared to conventional CNNs.

In "CorFormer: A Hybrid Transformer-CNN Architecture for Corrosion Segmentation on Metallic Surfaces", the authors present CorFormer, a real-time segmentation model that combines CNNs and Transformers within a unified framework [47]. Transformer layers are embedded within the CNN encoder, while a Semantic Gap Merger (SGM) bridges feature-level disparities. A hierarchical decoder further processes multi-scale features, enabling the model to capture both localized and broad defects. CorFormer improved Intersection-over-Union (IoU) by 2.7% over baseline methods and achieved real-time performance at 28 FPS, making it well-suited for industrial monitoring applications.

The study "Civil Infrastructure Damage and Corrosion Detection: An Application of Machine Learning" introduces a hybrid CycleGAN-based model incorporating Transformer-enhanced residual blocks for corrosion segmentation in UAV-acquired bridge imagery [48]. Positioned between the encoder and decoder of the generator, the Transformer blocks enable domain-adaptive learning, improving the model's robustness to texture and lighting variations. Trained on 1,300 annotated images from Melbourne's Bolte Bridge and Sky Rail, the model achieved an F1 score of 0.83 and an overall accuracy of 98.9%, outperforming traditional networks like PSPNet and SegNet, and demonstrating strong generalization capabilities across diverse aerial scenes.

The study "A Multilevel Bridge Corrosion Detection Method by Transformer-Based Segmentation in a Stitched View" addresses the limitations of CNNs when applied to wide, high-resolution structural images [49]. Utilizing a Transformer-based semantic segmentation model (e.g., SegFormer), the study incorporates stitched bridge surface imagery to enable large-area defect detection. Trained on 440 annotated images from Virginia Tech and additional data from Norway's Lundamo Bridge, the model achieved an F1 score of 68.2%, significantly outperforming U-Net and DeepLabV3+. Although performance was impacted by input resolution, the approach showed strong potential for deployment in field inspection workflows.

In "P-DETR: A Transformer-Based Algorithm for Pipeline Structure Detection", the authors propose an enhanced DETR architecture tailored for small-object detection in aerial pipeline imagery [50]. The model integrates a Feature Normalization and Transformation (FNT) module to improve spatial resolution and feature fusion. Trained on an augmented dataset of 6,127 drone-captured images, P-DETR achieved a mean Average Precision (mAP) of 55% and a recall of 73.9%, outperforming DETR, YOLOv3, and SSD baselines. While highly accurate in static image analysis, real-time drone-based deployment is identified as a future direction for further validation.

The study "Corrosion Segmentation Method of Concrete Drainage Pipes Based on Point Transformer" introduces a 3D point cloud segmentation method using a Point Transformer to identify corrosion in concrete drainage pipes [51]. A U-Net-like architecture is employed, trained on a combined dataset of real and simulated point clouds generated using an Azure Kinect DK. The model achieved 94.36% accuracy and a mean IoU of 86.31%, surpassing PointNet++ and PAConv. The best results were obtained with a 1:1 real-to-simulated data ratio. Despite its computational intensity, the model excels in precise 3D segmentation tasks for corrosion detection.

"SegFormer: Semantic Segmentation Based Transformers for Corrosion Detection" explores the use of SegFormer for corrosion classification based on annotated bridge inspection images following AASHTO and BIRM guidelines [52]. The dataset includes four corrosion severity levels, with class imbalance noted for the "severe" category. After fine-tuning, the model achieved a mean accuracy of 81.39% and a mean IoU of 71.16%. The study emphasizes the need for balanced datasets and tailored preprocessing to improve segmentation performance in underrepresented classes.

The paper "Double-Attention YOLO: Vision Transformer Model Based on Image Processing Technology in Complex Environment of Transmission Line Connection Fittings and Rust Detection" presents a two-stage framework for corrosion detection under visual degradation [53]. The approach integrates defogging preprocessing with a YOLO-based detector that combines Vision Transformers, a dual attention mechanism (CBAM), and GhostNet for efficient inference. Trained on 7,700 annotated images, the model achieved a mAP@0.5 of 99.48% and macro-F1 scores of 0.661 and 0.568. It outperformed multiple baselines—including ATSS, FCOS, and DETR variants—particularly in scenes with small objects, haze, and occlusion.

"Corrosion SAM: Adapting Segment Anything Model with Parameter-Efficient Fine-Tuning for Structural Corrosion Inspection" applies the SAM foundation model with ViT-H backbone to the corrosion domain using adapter-based Parameter-Efficient Fine-Tuning (PEFT) [54]. By inserting adapters into frozen Transformer layers, the model adapts with minimal computational overhead. Evaluated on the VDOT and KRDB datasets, Corrosion SAM achieved IoU scores of 69.72% and 60.78%, and Dice scores of 81.04% and 72.97%, respectively. It consistently outperformed DeepLabV3+, U-Net, and SegFormer, demonstrating the viability of efficient fine-tuning techniques in adapting foundation models to specialized inspection tasks.

Finally, "Pipeline Defect Recognition Using a CNN–Transformer Model" presents a lightweight architecture combining CNNs with Transformers and a Squeeze-and-Excitation module to classify defects in polyethylene (PE) gas pipelines [55]. The model omits positional encoding—leveraging zero-padding from CNNs—and uses Layer Normalization and multi-head attention for stability. Trained on 1,200 images captured by pipeline inspection robots, the model achieved an average accuracy of 97.8% over 10 randomized test runs. It consistently outperformed standard CNN and ViT architectures on small and medium-sized datasets, underscoring its potential for real-world pipeline monitoring.

As can be seen from these studies, the use of transformer-based models improved the accuracy of detecting anomalies caused by corrosion. Therefore, in this study, a transformer-based model was used on the experimental data set along with other models for a comparative analysis of the performances of each model.

IV. METHODOLOGY

A. Experimental Setup

The dataset used in this study was generated through a controlled laboratory experiment designed to simulate real-world corrosion in fluid transport pipelines. As shown in Fig. 1, ten acoustic emission (AE) sensors were strategically distributed along the pipeline—five on the left side (with girth welds between the source and sensors) and five on the right (without girth welds). Accelerated corrosion was induced at the centre of the pipeline over a three-hour period, with sensors spaced at 100 cm intervals to capture both near-field and far-field AE signal variations.

Fig. 1. Setup for the data acquisition.

This experimental dataset was chosen over public datasets to ensure high-fidelity, labelled AE signals under known and controlled corrosion conditions. The setup allowed us to evaluate the performance of machine learning models in detecting corrosion-induced anomalies while accounting for attenuation effects due to girth welds. Such a controlled environment also provided reliable ground truth for model validation, making the dataset particularly suitable for the supervised, unsupervised, and hybrid models evaluated in this study.

B. Data Analysis

After filtering and analysing the data, strong correlations was found between the parameters ABS-ENERGY and Signal Strength, along with distance (channel 3 being the closest and channel 8 being the furthest). The scatterplot in Fig. 2 shows the relationship. These relationships will be used to train the machine learning models that would detect the corrosion.

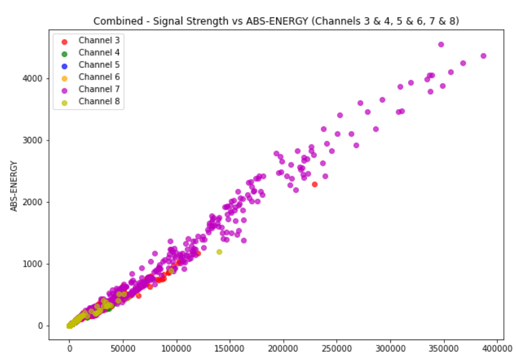

Fig. 2. Relationship between ABS-ENERGY and RMS.

V. RESULTS AND DISCUSSION

A. Machine Learning Models and Transformer Based Models Applied

In the following paragraphs, each ML model used, including the Hybrid CNN-Transformer, is described in detail along with its performance on the acquired data from the experiment. The bar graph in Fig. 3 provide a visual representation of the performances of each ML model.

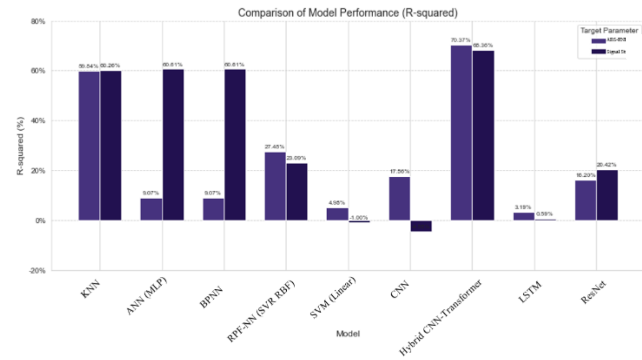

Fig. 3. Comparison of ML models.

1) K-Nearest Neighbours (KNN) is a non-parametric, instance-based learning algorithm that predicts the value of a target variable by averaging the values of its k-nearest neighbors in the feature space. For both 'ASL' and 'RMS', KNN demonstrated relatively strong performance, achieving R-squared values of 0.5984 and 0.6026, respectively. This suggests that the local patterns in the 'Distance' and 'CH' feature space are somewhat indicative of the target parameters. The moderate to high R-squared scores imply that a significant portion of the variance in both 'ASL' and 'RMS' can be explained by the values of their nearest neighbours. However, the performance is not perfect, indicating that other factors or more complex relationships might be at play.

2) The Artificial Neural Network (ANN) and Backpropagation Neural Network BPNN, both implemented as a Multi-Layer Perceptron with similar architectures (two hidden layers with 64 and 32 neurons), exhibited substantially lower performance for 'ASL' (R-squared of 0.0907) compared to KNN. For 'RMS', the performance was significantly better, reaching an R-squared of 0.6061, comparable to KNN. The discrepancy in performance for 'ASL' suggests that the non-linear relationships captured by the MLP were not as effective as the local averaging performed by KNN for this specific parameter. Conversely, for 'RMS', the MLP was able to model the underlying relationships as effectively as KNN. The fact that both ANN and BPNN yielded identical results is expected given their identical implementation and random initialization (controlled by the random state). This highlights the sensitivity of neural network performance to the specific target variable and the potential need for more extensive hyperparameter tuning.

3) The Support Vector Regression with Radial Basis Function kernel (RBF-NN), implemented using SVR with an RBF kernel, showed moderate predictive power for both 'ASL' (R-squared of 0.2748) and 'RMS' (R-squared of 0.2309). The RBF kernel allows the model to capture non-linear relationships through the use of radial basis functions. The performance was lower than KNN and the better performing neural network for 'RMS', indicating that the specific non-linear mapping learned by the RBF kernel was not as well-suited to the underlying data distribution for these parameters compared to the other effective models.

4) The Support Vector Regression (SVM) with a linear kernel demonstrated the poorest performance among the dedicated regression models for 'ASL', achieving an R-squared of only 0.0498. For 'RMS', the R-squared was even slightly negative (-0.0100), indicating that the model performed worse than simply predicting the mean of the 'RMS' values. This suggests that the relationship between the features ('Distance' and 'CH') and the target parameters is likely non-linear, and a linear model is insufficient to capture the underlying patterns. The near-zero or negative R-squared values highlight the importance of selecting a model that can accommodate the complexity of the data relationships.

5) The Convolutional Neural Network (CNN) adapted for this task by treating each data point as a single timestep, also exhibited low performance, with an R-squared of 0.1756 for 'ASL' and a negative R-squared of -0.0455 for 'RMS'. CNNs are typically designed for sequential or spatial data, and their poor performance here likely stems from the lack of a meaningful sequential structure in the provided features in relation to the target variables. The convolutional layers, designed to detect local patterns across a sequence, may not have been effectively utilized in this context.

6) The Hybrid CNN-Transformer Model yielded the highest performance for both 'ASL' (R-squared of 0.7037) and 'RMS' (R-squared of 0.6836) among all the models evaluated. This model leverages the strengths of Convolutional Neural Networks (CNNs) for local feature extraction from the raw acoustic emission (AE) waveforms or their spectrograms, and Transformer encoders for capturing long-range temporal dependencies. The CNN layers are adept at identifying intricate, localized patterns and transient events within the AE signals, which are crucial for understanding the underlying physical processes. Subsequently, the Transformer's self-attention mechanism processes these extracted features, allowing the model to learn complex, non-linear relationships across extended segments of the AE data. The high R-squared scores suggest that the model effectively translates the rich information embedded in the AE signals (which implicitly capture 'Distance' and 'CH' effects, along with other nuanced characteristics) into highly accurate predictions for both 'ASL' and 'RMS'. The strong performance of this hybrid architecture indicates its superior ability to learn hierarchical and contextual representations from the sequential AE data, bridging the gap between local signal characteristics and their global impact on the target parameters.

7) The Long Short-Term Memory network (LSTM), a type of recurrent neural network designed for sequential data, showed very poor performance, with an R-squared of 0.0319 for 'ASL' and 0.0059 for 'RMS'. Similar to the CNN, the poor performance is likely due to the lack of a clear temporal dependency in the provided features. Treating the data points as a sequence of single timesteps did not allow the LSTM's recurrent architecture to effectively learn predictive patterns for these parameters.

8) The Deep Residual Network (ResNet) implemented also showed relatively low performance, with an R-squared of 0.1620 for 'ASL' and 0.2042 for 'RMS'. While ResNets are powerful for learning deep representations, their architecture, adapted here for non-sequential data, did not outperform simpler models like KNN or the Decision Tree. The skip connections, designed to ease the training of deep networks, might not have provided a significant advantage in this context with the given data structure.

VI. Limitations and Future Work

While the study demonstrates the potential of machine learning and Transformer-based models for corrosion detection using acoustic emission data, several limitations should be acknowledged. One notable limitation is that the effect of temperature on the pipeline and the acoustic signal propagation was not considered during the experiment. In real-world scenarios, pipelines operate under varying thermal conditions, which can alter AE signal characteristics such as amplitude, attenuation, and wave speed. Ignoring temperature effects may limit the generalizability of the trained models to operational environments with fluctuating temperatures.

Future work should incorporate temperature as a variable in both data collection and model training to improve robustness. This could involve using temperature-compensated sensors, collecting data under controlled thermal variations, or integrating temperature data as an additional input feature for model training. Additionally, further research could explore the deployment of the proposed hybrid CNN-Transformer model in real-world pipelines with dynamic environmental conditions and test its performance across different corrosion types and pipeline materials.

VII. Conclusion

The study conclusively demonstrates the significant advantage of employing a Hybrid CNN-Transformer Model for accurately predicting Acoustic Signal Level (ASL) and Root Mean Square (RMS) in acoustic emission-based corrosion detection. Achieving R-squared values of 0.7037 for ASL and 0.6836 for RMS, this transformer-based architecture outperformed all other evaluated machine learning models, including traditional methods like KNN, SVM, and even other deep learning models like ANNs, CNNs, LSTMs, and ResNets.

This success can be attributed to the inherent design of the Hybrid CNN-Transformer. It effectively harnesses the Convolutional Neural Network's ability to extract nuanced, local features and transient patterns from raw AE waveforms or their spectrograms, which are critical for characterizing the distinct acoustic signatures of corrosion. Subsequently, the Transformer encoder's self-attention mechanism excels at learning complex, long-range temporal dependencies and contextual relationships across these extracted features. This synergistic combination allows the model to process the high-dimensional, sequential AE data comprehensively, bridging the gap between subtle local signal changes and their broader implications for corrosion-related parameters. The high R-squared values unequivocally indicate that this advanced, transformer-based approach can effectively decipher the intricate link between AE signals and the target parameters, representing a significant step forward in robust and accurate corrosion detection through acoustic emission monitoring.

ACKNOWLEDGMENT

This research is ongoing research supported by YUTP Research Grant Scheme.

REFERENCES

[1] N. R. Prasad, S. Almanza-Garcia, and T. T. Lu, “Anomaly detection,” Computers, Materials and Continua, vol. 14, no. 1, pp. 1–22, 2009, doi: 10.1145/1541880.1541882.

[2] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, “A review of novelty detection,” Jun. 2014. doi: 10.1016/j.sigpro.2013.12.026.

[3] D. Kim and T. Y. Heo, “Anomaly Detection with Feature Extraction Based on Machine Learning Using Hydraulic System IoT Sensor Data,” Sensors, vol. 22, no. 7, Apr. 2022, doi: 10.3390/s22072479.

[4] R. W. Anwar, M. Abrar, A. Salam, and F. Ullah, “Federated learning with LSTM for intrusion detection in IoT-based wireless sensor networks: a multi-dataset analysis,” PeerJ Comput Sci, vol. 11, pp. 1–30, 2025, doi: 10.7717/peerj-cs.2751.

[5] I. Priyadarshini, “Anomaly Detection of IoT Cyberattacks in Smart Cities Using Federated Learning and Split Learning,” Big Data and Cognitive Computing, vol. 8, no. 3, Mar. 2024, doi: 10.3390/bdcc8030021.

[6] J. C. Quiroz, N. Mariun, M. R. Mehrjou, M. Izadi, N. Misron, and M. A. Mohd Radzi, “Fault detection of broken rotor bar in LS-PMSM using random forests,” Measurement (Lond), vol. 116, pp. 273–280, Feb. 2018, doi: 10.1016/j.measurement.2017.11.004.

[7] F. Ferraz Júnior, R. A. F. Romero, and S. J. Hsieh, “Machine Learning for the Detection and Diagnosis of Anomalies in Applications Driven by Electric Motors,” Sensors, vol. 23, no. 24, Dec. 2023, doi: 10.3390/s23249725.

[8] J. Garcia, L. Rios-Colque, A. Peña, and L. Rojas, “Condition Monitoring and Predictive Maintenance in Industrial Equipment: An NLP-Assisted Review of Signal Processing, Hybrid Models, and Implementation Challenges,” May 01, 2025, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/app15105465.

[9] R. Abdalla, H. Samara, N. Perozo, C. P. Carvajal, and P. Jaeger, “Machine Learning Approach for Predictive Maintenance of the Electrical Submersible Pumps (ESPs),” ACS Omega, vol. 7, no. 21, pp. 17641–17651, May 2022, doi: 10.1021/acsomega.1c05881.

[10] L. Yu, E. Yang, C. Luo, and P. Ren, “AMCD: an accurate deep learning-based metallic corrosion detector for MAV-based real-time visual inspection,” J Ambient Intell Humaniz Comput, vol. 14, no. 7, pp. 8087–8098, Jul. 2023, doi: 10.1007/S12652-021-03580-4/FIGURES/12.

[11] D. J. Atha and M. R. Jahanshahi, “Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection,” Struct Health Monit, vol. 17, no. 5, pp. 1110–1128, Sep. 2018, doi: 10.1177/1475921717737051;CTYPE:STRING:JOURNAL.

[12] W. Wu, T. Tao, J. Xiao, Y. Yao, and J. Yang, “Unsupervised Anomaly Detection on Metal Surfaces Based on Frequency Domain Information Fusion,” Sensors 2025, Vol. 25, Page 2250, vol. 25, no. 7, p. 2250, Apr. 2025, doi: 10.3390/S25072250.

[13] T. Okura, N. Kasai, H. Minowa, and S. Okazaki, “Application of Machine Learning for Data with an Atmospheric Corrosion Monitoring Sensor Based on Strain Measurements,” Metals 2022, Vol. 12, Page 1179, vol. 12, no. 7, p. 1179, Jul. 2022, doi: 10.3390/MET12071179.

[14] M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander, “LOF: Identifying Density-Based Local Outliers,” SIGMOD 2000 - Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp. 93–104, 2000, doi: 10.1145/342009.335388.

[15] F. T. Liu, K. M. Ting, and Z. H. Zhou, “Isolation-based anomaly detection,” ACM Trans Knowl Discov Data, vol. 6, no. 1, Mar. 2012, doi: 10.1145/2133360.2133363.

[16] E. Calikus, S. Nowaczyk, and O. Dikmen, “Context discovery for anomaly detection,” Int J Data Sci Anal, vol. 19, no. 1, pp. 99–113, Jan. 2024, doi: 10.1007/S41060-024-00586-X/FIGURES/5.

[17] M. A. Hayes and M. A. Capretz, “Contextual anomaly detection framework for big sensor data,” J Big Data, vol. 2, no. 1, pp. 1–22, Dec. 2015, doi: 10.1186/S40537-014-0011-Y/TABLES/9.

[18] L. Bontemps, V. L. Cao, J. McDermott, and N. A. Le-Khac, “Collective Anomaly Detection based on Long Short Term Memory Recurrent Neural Network,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10018 LNCS, pp. 141–152, Mar. 2017, doi: 10.1007/978-3-319-48057-2_9.

[19] D. M. Hawkins, “Identification of Outliers,” Identification of Outliers, 1980, doi: 10.1007/978-94-015-3994-4/COVER.

[20] D. C. Montgomery, “Introduction to Statistical Quality Control, 8th Edition | Wiley,” Wiley Online Library, p. 786, 2019, Accessed: Jun. 24, 2025. [Online]. Available: https://www.wiley.com/en-us/Introduction+to+Statistical+Quality+Control%2C+8th+Edition-p-9781119399308

[21] F. E. Grubbs, “AD-781 499 PROCEDURES FOR DETECTING OUTLYING OBSERVATIONS IN SAMPLES,” 1974.

[22] F. Gustafsson, “Adaptive Filtering and Change Detection,” 2000, Accessed: Jun. 24, 2025. [Online]. Available: http://www.wiley.co.uk

[23] “Hyndman, R.J. and Athanasopoulos, G. (2018) Forecasting Principles and Practice. OTexts. - References - Scientific Research Publishing.” Accessed: Jun. 24, 2025. [Online]. Available: https://www.scirp.org/reference/referencespapers?referenceid=2849375

[24] M. Lavielle, “Using penalized contrasts for the change-point problem,” Signal Processing, vol. 85, no. 8, pp. 1501–1510, Aug. 2005, doi: 10.1016/J.SIGPRO.2005.01.012.

[25] R. Killick, P. Fearnhead, and I. A. Eckley, “Optimal detection of changepoints with a linear computational cost,” J Am Stat Assoc, vol. 107, no. 500, pp. 1590–1598, 2012, doi: 10.1080/01621459.2012.737745.

[26] C.-C. M. Yeh et al., “Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets,” pp. 1317–1322, Feb. 2017, doi: 10.1109/ICDM.2016.0179.

[27] J. Antoni, “The spectral kurtosis: a useful tool for characterising non-stationary signals,” Mech Syst Signal Process, vol. 20, no. 2, pp. 282–307, Feb. 2006, doi: 10.1016/J.YMSSP.2004.09.001.

[28] “A Wavelet Tour of Signal Processing | ScienceDirect.” Accessed: Jun. 24, 2025. [Online]. Available: https://www.sciencedirect.com/book/9780123743701/a-wavelet-tour-of-signal-processing

[29] R. B. Randall, “Vibration-Based Condition Monitoring: Industrial, Aerospace and Automotive Applications,” Vibration-Based Condition Monitoring: Industrial, Aerospace and Automotive Applications, pp. 1–289, Dec. 2010, doi: 10.1002/9780470977668.

[30] C. U. Grosse and M. Ohtsu, “Acoustic emission testing: Basics for Research-Applications in Civil Engineering,” Acoustic Emission Testing: Basics for Research-Applications in Civil Engineering, pp. 1–404, 2008, doi: 10.1007/978-3-540-69972-9/COVER.

[31] M. Ahmed, A. Naser Mahmood, and J. Hu, “A survey of network anomaly detection techniques,” Journal of Network and Computer Applications, vol. 60, pp. 19–31, Jan. 2016, doi: 10.1016/J.JNCA.2015.11.016.

[32] J. Ma and S. Perkins, “Time-series Novelty Detection Using One-class Support Vector Machines,” Proceedings of the International Joint Conference on Neural Networks, vol. 3, pp. 1741–1745, 2003, doi: 10.1109/IJCNN.2003.1223670.

[33] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, “High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning,” Pattern Recognit, vol. 58, pp. 121–134, Oct. 2016, doi: 10.1016/J.PATCOG.2016.03.028.

[34] L. Jing and Y. Tian, "Self-supervised Visual Feature Learning with Deep Neural Networks: A Survey," IEEE Trans Pattern Anal Mach Intell, vol. 43, no. 11, pp. 4037–4058, Feb. 2019, doi: 10.1109/TPAMI.2020.2992393.

[35] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. Part F129685, pp. 665–674, Aug. 2017, doi: 10.1145/3097983.3098052/SUPPL_FILE/ZHOU_ANOMALY_DETECTION.MP4.

[36] J.-Y. Franceschi, A. Dieuleveut, and M. Jaggi, "Unsupervised Scalable Representation Learning for Multivariate Time Series," Adv Neural Inf Process Syst, vol. 32, 2019.

[37] C. Ding and C. Wu, "Self-Supervised Learning for Biomedical Signal Processing: A Systematic Review on ECG and PPG Signals," medRxiv, p. 2024.09.30.24314588, Oct. 2024, doi: 10.1101/2024.09.30.24314588.

[38] "Long Short Term Memory Networks for Anomaly Detection in Time Series." Accessed: Jun. 24, 2025. [Online]. Available: https://www.researchgate.net/publication/304782562_Long_Short_Term_Memory_Networks_for_Anomaly_Detection_in_Time_Series

[39] J. An and S. Cho, "SNU Data Mining Center 2015-2 Special Lecture on IE Variational Autoencoder based Anomaly Detection using Reconstruction Probability," 2015.

[40] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10265 LNCS, pp. 146–147, Mar. 2017, doi: 10.1007/978-3-319-59050-9_12.

[41] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 387–395, Jul. 2018, doi: 10.1145/3219819.3219845/SUPPL_FILE/HUNDMAN_DETECTING_SPACECRAFT_ANOMALIES.MP4.

[42] A. Vaswani et al., "Attention Is All You Need," Aug. 2023, [Online]. Available: http://arxiv.org/abs/1706.03762

[43] S. Li et al., "Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting," Adv Neural Inf Process Syst, vol. 32, Jun. 2019, Accessed: Jun. 24, 2025. [Online]. Available: https://arxiv.org/pdf/1907.00235

[44] Z. Rong, D. Shandong, N. Xin, and X. Shiguang, "Feedforward Neural Network for Time Series Anomaly Detection," Dec. 2018, Accessed: Jun. 24, 2025. [Online]. Available: https://arxiv.org/pdf/1812.08389

[45] "Sensors | Special Issue : Structural Health Monitoring Using Acoustic Emission Signals: The Role of Machine Learning." Accessed: Jun. 24, 2025. [Online]. Available: https://www.mdpi.com/journal/sensors/special_issues/QE6G26RX6Y

[46] J. Guo, L. Wang, and L. Hua, "Efficient Metal Corrosion Area Detection Model Combining Convolution and Transformer," Applied Sciences 2024, Vol. 14, Page 9900, vol. 14, no. 21, p. 9900, Oct. 2024, doi: 10.3390/APP14219900.

[47] A. Subedi, C. Qian, R. Sadeghian, and M. R. Jahanshahi, "CorFormer: a hybrid transformer-CNN architecture for corrosion segmentation on metallic surfaces," Mach Vis Appl, vol. 36, no. 2, pp. 1–21, Mar. 2025, doi: 10.1007/S00138-025-01663-2/METRICS.

[48] H. S. Munawar, F. Ullah, D. Shahzad, A. Heravi, S. Qayyum, and J. Akram, "Civil Infrastructure Damage and Corrosion Detection: An Application of Machine Learning," Buildings 2022, Vol. 12, Page 156, vol. 12, no. 2, p. 156, Feb. 2022, doi: 10.3390/BUILDINGS12020156.

[49] Z. Lu, T. Jiang, J. Slavič, · Gunnstein, and T. Frøseth, "A multilevel bridge corrosion detection method by transformer-based segmentation in a stitched view," J Civ Struct Health Monit, 123AD, doi: 10.1007/s13349-025-00928-8.

[50] I. A. Aromoye, L. H. Hiung, and P. Sebastian, "P-DETR: A transformer-based algorithm for pipeline structure detection," Results in Engineering, vol. 26, p. 104652, Jun. 2025, doi: 10.1016/J.RINENG.2025.104652.

[51] N. Wang et al., "Corrosion segmentation method of concrete drainage pipes based on point transformer," Measurement, vol. 234, p. 114893, Jul. 2024, doi: 10.1016/J.MEASUREMENT.2024.114893.

[52] A. Safa, A. Mohamed, B. Issam, and H. Mohamed-Yassine, "SegFormer: Semantic Segmentation Based Tranformers for Corrosion Detection," 6th International Conference on Networking and Advanced

[53] Z. Song, X. Huang, C. Ji, and Y. Zhang, "Double-Attention YOLO: Vision Transformer Model Based on Image Processing Technology in Complex Environment of Transmission Line Connection Fittings and Rust Detection," Machines 2022, Vol. 10, Page 1002, vol. 10, no. 11, p. 1002, Oct. 2022, doi: 10.3390/MACHINES10111002.

[54] C. Chai, Y. Gao, H. Li, and X. Zhu, "Corrosion SAM: Adapting Segment Anything Model with Parameter-Efficient Fine-Tuning for Structural Corrosion Inspection".

[55] X. Niu, X. Guo, H. Song, W. Hu, G. Jiang, and J. Guan, "Pipeline Defect Recognition Algorithm Based on CNN Transformer Model for Internal Detection in PE Gas Pipelines," Proceedings of 2024 Academic Conference of China Instrument and Control Society, ACCIS 2024, pp. 531–536, 2024, doi: 10.1109/ACCIS62068.2024.10948683.