META: Deep Learning Pipeline for Detecting Anomalies on Multimodal Vibration Sewage Treatment Plant Data

Simeon Krastev[1], Aukkawut Ammartayakun[1], Kewal Jayshankar Mishra[1], Harika Koduri[1], Eric Schuman[1], Drew Morris[1], Yuan Feng[1], Sai Supreeth Reddy Bandi[1], Chun-Kit Ngan[1], Andrew Yeung[2], Jason Li[2], Nigel Ko[2], Fatemeh Emdad[1], Elke Rundensteiner[1], Heiton M. H. Ho[3], T. K. Wong[3] and Jolly P. C. Chan[3]

[1]*Data Science Program, Worcester Polytechnic Institute, 100 Institute Rd., Worcester, MA, U.S.A.*

[2]*XTRA Sensing Limited, Wan Chai, Hong Kong Island, Hong Kong SAR*

[3]*Drainage Services Department, The Government of the Hong Kong Special Administrative Region, Wanchai, Hong Kong Island, Hong Kong SAR*

{*sdkrastev, aammartayakun, kmishra1, hkoduri, demorris, erschuman, yfeng8, sbandi, cngan, femdad, rundenst*}*@wpi.edu,*

Keywords: Predictive Maintenance, Anomaly Detection, Signal Averaging, Data Fusion, Multimodal Feature Extraction, Autoencoder, Transformer Model, Sewage Treatment Plants.

Abstract: In this paper, we propose a hybrid anomaly detection pipeline, META, which integrates Multimodal-feature Extraction (ME) and a Transformer-based Autoencoder (TA) for predictive maintenance of sewage treatment plants. META uses a three-step approach: First, it employs a signal averaging method to remove noise and improve the quality of signals related to pump health. Second, it extracts key signal properties from three vibration directions (Axial, Radial X, Radial Y), fuses them, and performs dimensionality reduction to create a refined PCA feature set. Third, a Transformer-based Autoencoder (TA) learns pump behavior from the PCA features to detect anomalies with high precision. We validate META with an experimental case study at the Stonecutters Island Sewage Treatment Works in Hong Kong, showing it outperforms state-of-the-art methods in metrics like MCC and F1-score. Lastly, we develop a web-based Sewage Pump Monitoring System hosting the META pipeline with an interactive interface for future use.

1 INTRODUCTION

A sewage treatment plant (STP) is a critical infrastructure designed to process wastewater from residential, commercial, and industrial sources. Its objective is to remove contaminants, ensuring the treated water is safe and suitable for reuse before releasing it back to the environment. This process is essential for protecting public health and preserving the natural environment by adhering to water quality standards.

STPs support a sustainable environment and positively impact human life in several ways. For example, advanced sewage treatment, such as Singapore's NEWater program, significantly reduces the risk of waterborne diseases (Lee and Tan, 2016). By converting treated wastewater into ultra-clean, high-quality water, Singapore meets up to 40% of its water demand with NEWater, demonstrating a commitment to public health and environmental sustainability. The Groundwater Replenishment System (GWRS) in Orange County, California, demonstrates the potential of STPs in supporting sustainable agriculture and conserving freshwater resources (Ormerod and Silvia, 2017). By purifying treated wastewater to potable standards, GWRS provides a reliable water supply, reducing dependence on imported water. In Namibia, Windhoek Goreangab Operating Company (WINGOC) provides treated wastewater which is extensively used for agricultural irrigation (Lahnsteiner and Lempert, 2007). This approach not only conserves freshwater resources but also supports sustainable agriculture by providing a reliable water source for irrigation.

The Stonecutters Island Sewage Treatment Works (SCISTW) is one of the most efficient chemical treatment plants in the world, removing 70% of the organic pollutants in terms of biochemical oxygen demand; 80% of the suspended solids; and 50% of sewage pathogens in terms of Escherichia coli (E.

Krastev, S., Ammartayakun, A., Mishra, K., Koduri, H., Schuman, E., Morris, D., Feng, Y., Bandi, S., Ngan, C., Yeung, A., Li, J., Ko, N., Emdad, F., Rundensteiner, E., Ho, H., Wong, T. and Chan, J.
META: Deep Learning Pipeline for Detecting Anomalies on Multimodal Vibration Sewage Treatment Plant Data.
DOI: 10.5220/0013031600003837
In *Proceedings of the 16th International Joint Conference on Computational Intelligence (IJCCI 2024)*, pages 461-474
ISBN: 978-989-758-721-4; ISSN: 2184-3236

Figure 1: Stonecutters Island Sewage Treatment Works is located in Victoria Harbour and collects sewage from the Hong Kong region for processing. The project is focused on the main pumping stations (3) that pump the filtered sewage up to the chemical processing tanks (Drainage Services Department, 2009b).

coli) level, which is an indicator of disease causing microorganisms. In addition, the levels of key pollutants in the harbour-area waters have generally decreased. Ammonia, which is toxic to marine life, declined by 25%; nutrients, in terms of total inorganic nitrogen and phosphorus, have reduced by 16% and 36%, respectively; and the overall E. coli level has reduced by around 50% (Drainage Services Department, 2009a).

The current strategy to prevent failures in large-scale industrial pumps like those found in STPs is to follow a routine maintenance schedule. This approach involves servicing the pumps at regular intervals and replacing parts that cannot be repaired over the life of the pump. For example, the bearings may be repacked with grease annually until they reach their rated life, at which point the entire bearing assembly must be replaced. This approach is relatively simple to implement and does prevent failures due to wear and age, but it is not able to capture failures due to incorrect installation or operation. Another issue with this approach is that parts may be prematurely replaced, which increases part cost and may unnecessarily bring about machine downtime (Carson, 2011).

Predictive maintenance, also referred to as condition-based maintenance, enhances routine maintenance by continuously monitoring the condition of system components, replacing those that exhibit signs of impending failure. A primary benefit of such an approach is that a broader range of causes of failure can be prevented. For example, an incorrectly installed component would start to show signs of wear that the predictive model could detect and signal a need for replacement despite the service life not being reached.

However, this approach requires a more complex system to predict and identify potential failures before they occur. A traditional approach for doing this usually requires the development of hard-coded rules and algorithms to define and then detect abnormal operation. This can easily grow exponentially in cost and complexity to implement (Newton, 2021); additionally, it requires prior knowledge and definition of all possible anomalies. An alternative approach, which this paper proposes, is to instead leverage state-of-the-art deep learning techniques and neural networks to automatically learn data-driven strategies that characterize abnormal operation (Faisal et al., 2023).

Such a technique requires the fusion of multiple sources of pump vibration data (Diez-Olivan et al., 2019). If intelligently fused, the anomaly detection model can then learn from a much broader space instead of just learning one source of vibration data. For example, fusing three directions of pump vibration data together may pick up anomalous behavior that does not arise in just one direction. In addition, the data dimensions can be reduced through modeling techniques such as Autoencoders and Principal Component Analysis (PCA). These techniques can remove noise and redundant information from the data streams, resulting in a cleaner signal for the anomaly detection models to work on.

Thereafter, the multi-modal preprocessed data can be used as the input to modern deep learning methods such as transformer models. Transformers learn complex possibly non-linear relationships in data through a deep neural network architecture and attention mechanisms (Vaswani et al., 2017), which focus on different aspects of the data to understand normal pump behavior. Combining this feature with an ability to track changes over time, the models can tell when the pump is deviating from normal activity.

In the literature, hybrid machine learning (ML) models and neural networks have been used to diagnose power transformers using acoustic signals (Yu et al., 2023), while transformer models have been utilized for fault detection and classification in manufacturing (Wu et al., 2023). However, research on the applications of these models remains limited. In addition, a few studies (Diez-Olivan et al., 2019) have explored leveraging multiple input data formats and combining data from various sources to enhance detection performance.

To leverage the strengths of these methodologies for predictive maintenance of STPs, we propose a innovative anomaly detection platform, called META, which integrates Multimodal-feature Extraction (ME)

and a Transformer-based Autoencoder (TA) solution. Our contributions to this work are five-fold:

1. We develop a signal averaging method to remove unrelated noise from the raw sensor data and improve the quality of signals related to the pump health and operations.
2. We extract the meaningful signal properties from three vibration signal directions (i.e., Axial, Radial X, and Radial Y) using multimodal-feature extraction methods. We then fuse these properties together and reduce data dimensionality using PCA to generate a refined PCA feature set.
3. We apply a Transformer-based Autoencoder (TA) model to learn pump behavior from the extracted PCA feature set to detect anomalous behavior.
4. We conduct an extensive experimental case study on the SCISTW, located in Hong Kong, in which we reconstruct the vibration signals of the pumps and set a threshold to detect anomalies based on reconstruction error. Having labeled data on the dates in which the pumps were operating abnormally, we are able to obtain accuracy metrics on the performance of the META pipeline. META achieves MCC/F1 scores of 0.966/0.995 in anomaly detection, a percentage increase of 4.89%/2.47% compared to the state-of-the-art anomaly detection performance by (Yu et al., 2023), and a percentage increase of 2.77%/3.977% compared to (Wu et al., 2023).
5. We create a web prototype for a Sewage Pump Monitoring System hosting the META pipeline, providing an interactive interface for future use.

The rest of the paper is organized as follows. In Section 2, we review the previous work and literature on anomaly detection methods and explore the existing data fusion techniques and transformer-based models in predictive maintenance. In Section 3, we provide an overview of the SCISTW. In Section 4, we introduce our META framework, broadly describing its core components. Starting in Section 4.1, we examine each of its components, beginning with the Signal Averaging method. In Section 4.2, we explain our ME approach. In Section 4.3, we present our TA Model, explaining its architecture and the mechanism that it uses to detect anomalies. In Section 5, we conduct an extensive experimental study and analyze the performance of the META platform in detecting anomalies at the SCISTW. In Section 6, we describe the development of our GUI-based application for monitoring sewage pumps and visualizing the analysis results. Finally, in Section 7, we conclude the paper and outline the potential future work to enhance our META framework's performance and explore its application in other types of industrial machinery.

2 RELATED WORK

The development of anomaly detection methods for rotating machinery is an active area. Three data fusion techniques are reviewed and explored for rotating machinery. In Yuan et al.'s work, a flexible framework is proposed to fuse multiple sources together to detect abnormal operation. Multiple 1-d and 2-d signals are first individually passed through convolutional neural networks and subsequently joined using t-distributed Stochastic Neighbor Embedding (tSNE) (Yuan et al., 2018). The benefit of this approach is the flexibility in its application which would allow it to be adapted to other domains. Pang et al.'s work proposes a model that applies Stacked Denoising AutoEncoders (SDAEs) to the individual data sources and fuses these sources together via Local and Global Principal Components Analysis (LGPCA) with the aim of improving the representation of the signal. The SDAEs remove the uncorrelated noise from the source signal and the LGPCA combines the signals together across different scales to ensure that the temporal relationships are preserved (Pang et al., 2020). Finally, in Wang et al.'s work, the authors propose a method of incorporating feature selection into the model training processing by computing the Mean Impact Value (MIV) and Within-class and Between-class Discriminant Analysis (WBDA) to dynamically select features. The approach uses traditional measures to assess the health of rotating machinery and aims to combine these features so that a greater representation and context is achieved while reducing data redundancy and noise (Wang et al., 2021).

Two transformer models are reviewed and explored in predictive maintenance of rotating machinery. In Yu et al.'s work, the authors propose a hybrid anomaly detection model that combines a convolutional neural network and a recurrent neural network with the attention mechanism found in a transformer. This approach uses the CNN to extract relevant patterns from the input signals and the recurrent layer to capture the long-running state of the system and feeds this information into the attention mechanism to detect anomalous operation (Yu et al., 2023). In Wu et al.'s work, the authors adapt a traditional transformer to work in the classification setting for the purpose of detecting abnormal operation of the pump. The transformer is used as an encoder to embed the input data and generate a classification by passing the embedding to a fully connected network (Wu et al., 2023).

Figure 2: The motor that drives one of the pumps in main pumping station 2.

3 BACKGROUND: STONECUTTERS ISLAND SEWAGE TREATMENT

The SCISTW in Hong Kong is a core infrastructure that makes every effort to minimize unplanned interruptions in service. It is one of the largest and the most compact sewage treatment works of its type in the world, occupying 10 hectares of reclaimed land, and is designed to provide Chemically Enhanced Primary Treatment for an average flow of 1.7 million cubic meters per day (Drainage Services Department, 2009a). The overall process of treating sewage is shown in Figure 1. In order to pump 1.7 million cubic meters of sewage per day, the SCISTW has two main pumping facilities that run continuously. Both pumping stations are comprised of eight sewage pumps that are able to be independently driven at a range of frequencies or completely shutdown in order to maintain a target flow rate into the treatment pools. Direct-drive pumps pump sewage approximately 40 meters up from the collection tunnels to the treatment pools. The AC induction motor used to drive the pump is shown in Figure 2; the main pump body is shown in Figure 3. The sewage pumps that move the raw

Figure 3: Main pump body that transfers sewage from the collection tunnels to the treatment pools 40 meters above.

sewage from the collection tunnels up to the treatment tanks are critical to the operation of the treatment plant. The current regime for failure prevention at the SCISTW is largely based on preventative maintenance, relying on the excess pumping capacity available in the plant by using working pumps to handle the load when pumps fail. Shifting to a condition-based maintenance approach would allow the SCISTW to prevent pumps from failing unexpectedly and enable the plant to treat a greater load of sewage as a result.

The primary hurdle to the SCISTW adopting a condition-based maintenance approach is the cost and complexity of implementing a hard-coded approach to the pumps. While the design of those pumps is relatively simple, the number of failure modes is high,

and the interaction between failure modes increases the complexity of creating logic that may not accurately detect when a pump is about to fail. Furthermore, the pumps are controlled and monitored through several disparate systems, making it difficult to create a holistic anomaly detection system that can integrate all the data coming from the different sources. As a solution, the SCISTW is looking towards implementing an independent monitoring system capable of detecting abnormalities autonomously, without the need to tap into each system individually.

4 META OVERVIEW: ANOMALY DETECTION PLATFORM

Our pipeline for anomaly detection incorporates a method of fusing multi-directional raw acceleration time waveform data (Axial, Radial_X, and Radial_Y) and using the resultant product as an input for a transformer model architecture. This pipeline, shown in Figure 4, consists of three core components, including Signal Averaging (SA), Multimodal-feature Extraction, and a Transformer-based Autoencoder Model.

The purpose of the SA module is to process the three time waveform signals (Axial, Radial_X, and Radial_Y) collected from the accelerometer. These signals are first split into time segments based on the shaft periods and then averaged to preserve the segment signals specific to each individual pump, effectively removing unrelated noise. After that, the cleaned signals are sent to the ME module that summarizes the specific aspects of the data by using statistical measures, empirical mode decomposition (EMD), entropy measures, wavelet packet decomposition (WPD), and frequency domain analysis (FDA), respectively. The summarized data of each time waveform signal are then further processed using Mean Impact Value (MIV) and Within-class and Between-class Discriminant Analysis (WBDA) for feature selection. The final selected feature set is then fused by PCA. Those PC features are combined for each time waveform signal and then further concatenated from all these three directions to form the final feature set that is then sent to the TA model for the anomaly detection. This model processes the final feature set through flattening, patching, and positional encoding layers, before going through a multi-head attention encoder. The output is then sent through several dense layers to reconstruct the signal. Once trained, this model architecture is used to classify pump behavior as normal or anomalous based on reconstruction error. Once the error threshold is determined, any major deviation from the reconstruction error of a model trained on normal data, i.e., above the error threshold, could be a potential anomaly.

4.1 META Signal Averaging Strategy

The data that is analyzed by the META framework to detect abnormal pump operations is collected from various locations on the pumps using a three-axis accelerometer. The data is collected in two seconds every four hours in all three directions, yielding 25,600 data points per two-second recording. The accelerometer is able to record all frequencies between 0.5 and 5,000 Hz. A two-second recording is long enough to record several full rotations of the pump so that pump health can be determined.

While the ability to record such a large range of signals has the benefit of being able to detect multiple types of failures like shaft imbalances at the low

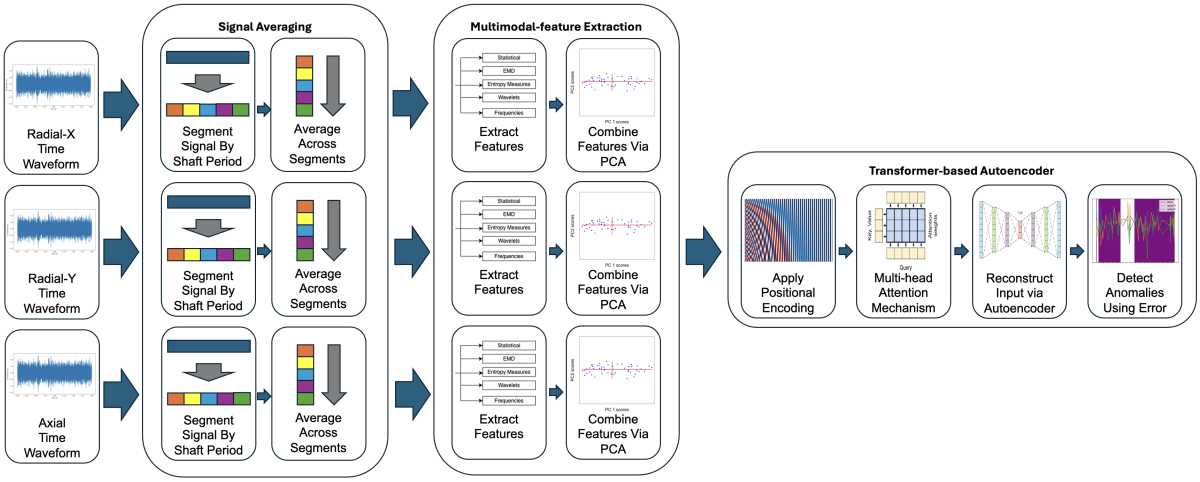

Figure 4: META Framework.

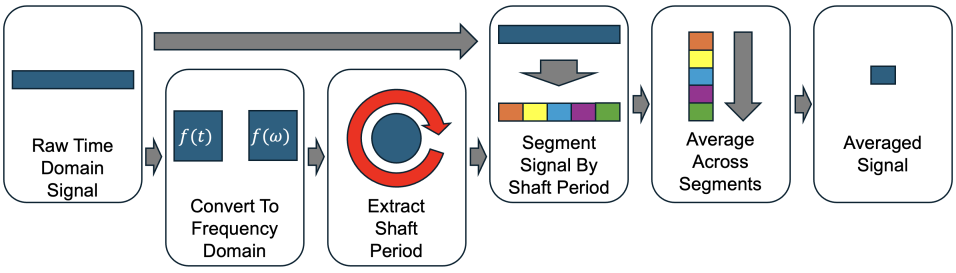

Figure 5: 1) The raw time-based signal is first used to determine the rotation period of the pump. 2) The extracted period is then used to segment the raw time-based signal into windows equal to the rotation period. This aligns the angle of rotation for elements across windows. 3) The segmented windows are then averaged across the segments to generate a new segment which should reduce the uncorrelated noise while preserving the relevant signals.

end and seized bearings at the high end, it also has the drawback of capturing a lot of information in the process. In fact, the system's broadband noise is so significant that it poses substantial challenges for many anomaly detection algorithms, making it difficult to distinguish between signals from a healthy pump and an unhealthy one. Our META framework can address this issue by incorporating signal averaging as a preprocessing step to improve the signal-to-noise ratio.

Signal averaging is applied to the raw time waveform signal in an attempt to improve the signal-to-noise ratio of the captured signal. At its core, the pumps are rotating machinery, and that means the majority of vibration signals relevant to the operation of the pump is periodic in nature. Furthermore, the short duration of the recording means the speed of the pumps can be assumed to be steady-state. Exploiting this fact allows the original time waveform signal to be split into time segments equal to the rotation period of the pump and then averaged together. Averaging these segments together is expected to preserve the signals specific to the pump because they "constructively" overlap. However, the random noise not related to the operation of the pump is expected to "destructively" overlap, as it is assumed that the noise has a mean of 0. By averaging multiple segments together, the signals related to the pump are elevated above the noise providing a cleaner signal to the work downstream. Signal averaging is implemented in our META framework shown in Figure 5.

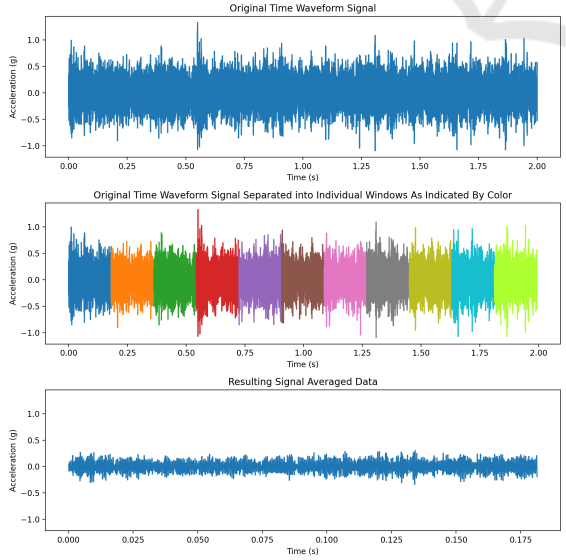

Figure 6: Top) The original time waveform data. Middle) The original time waveform data segmented into windows of time that match the period of rotation. Bottom) Resulting signal averaged data.

Applying signal averaging as a preprocessing step makes the assumption that the noise that is uncorrelated to the rotation is of less importance than the signals and the noise that is correlated to the rotation. The rotation of the shaft is the sole energizing component of the system, making this a safe assumption. Furthermore, while signal averaging reduces the noise floor of the resulting signal, it does not remove the noise from the signal altogether. This provides some protection against the assumption proving to be false. One notable issue with applying signal averaging to the data is that it may greatly diminish transient signals that do not repeatedly appear across multiple rotations of the pump. An example of passing the data through signal averaging is shown in Figure 6.

4.2 META Multimodal Feature Extraction

In the context of enhancing anomaly detection mechanisms for pump systems, the presented method adopts a multimodal feature extraction approach, leveraging the tri-axial nature of pump operation—Radial_X, Radial_Y, and Axial. This approach facilitates a comprehensive analysis by considering the acceleration data across each dimension, enabling the capture of a full spectrum of potential anomalies. The engineering process produces an initial robust set of features in-

corporating time-domain statistics, frequency-domain characteristics, and time-frequency domain features. These features, which are meticulously named to reflect their source and extraction method, shown in Figure 7, range from basic statistical measures, such as mean, standard deviation, skewness, and kurtosis, to more complex analyses including empirical mode decomposition (EMD) (i.e., energy and entropy), entropy measures (i.e., permutation and dispersion), wavelet packet decomposition, and frequency measures (i.e., sum amplitude, average spectrum, standard deviation spectrum, and integral spectrum).

The ME approach therefore offers a comprehensive analysis of the pump system's operational data. The concatenation of these features enables a robust anomaly detection system that can accurately detect and characterize a wide range of anomalous behaviors unique to each axis of operation, thereby enabling more reliable predictive maintenance. Additionally, it caters to the detection of complex, non-linear interactions within pump mechanics, thereby broadening the scope of detectable anomalies.

Following the feature selection process by MIV and WBDA, PCA is applied to each directional feature set independently. PCA, as a dimensionality reduction technique, identifies the principal components that encapsulate the maximum variance within the data, thus preserving essential information while reducing the data's complexity.

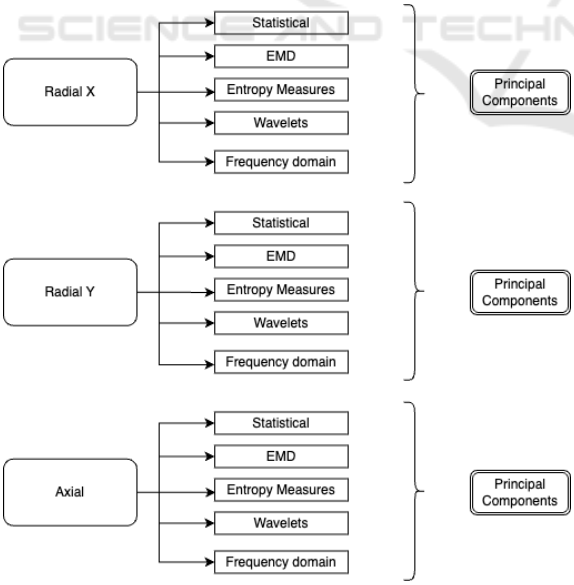

Figure 7: Multimodal-Feature Extraction.

For the PCA method explanation, we use Radial_X as an example. When PCA is applied to this signal, the process involves the following steps:

Standardization

The features extracted by each method from the Radial_X, shown in Figure 7, are standardized. This involves subtracting the mean of each feature and then dividing by the standard deviation. Standardization ensures that each feature contributes equally to the analysis.

$$\mathbf{X}_{\text{Radial\_X, std}} = \frac{\mathbf{X}_{\text{Radial\_X}} - \mu_{\mathbf{X}_{\text{Radial\_X}}}}{\sigma_{\mathbf{X}_{\text{Radial\_X}}}} \quad (1)$$

Covariance Matrix Calculation

The covariance matrix of the standardized features is computed. The covariance matrix provides a measure of how much the features vary together.

$$\Sigma_{\mathbf{X}_{\text{Radial\_X}}} = \frac{1}{n-1}\mathbf{X}^{\top}_{\text{Radial\_X, std}}\mathbf{X}_{\text{Radial\_X, std}} \quad (2)$$

Eigenvalue and Eigenvector Computation

The eigenvalue equation is solved to obtain the eigenvalues and eigenvectors of the covariance matrix. The eigenvalues represent the amount of variance explained by each principal component, and the eigenvectors represent the direction of the principal components.

$$\Sigma_{\mathbf{X}_{\text{Radial\_X}}}\mathbf{v}_i = \lambda_i \mathbf{v}_i \quad (3)$$

Principal Component Selection

The top k principal components are selected by the eigenvectors corresponding to the largest eigenvalues. This step reduces the dimensionality of the data while retaining the most important information.

$$\mathbf{V}_k = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k] \quad (4)$$

Data Transformation

The standardized Radial_X data is transformed into the principal component space. This transformation projects the original data onto the new basis formed by the selected principal components.

$$\mathbf{X}_{\text{Radial\_X, PCA}} = \mathbf{X}_{\text{Radial\_X, std}}\mathbf{V}_k \quad (5)$$

Principal Components Concatenation

The same process is applied to the Radial_Y and Axial signals, resulting in the transformed data matrices $\mathbf{X}_{\text{Radial\_Y, PCA}}$ and $\mathbf{X}_{\text{Axial, PCA}}$. The principal components from all three directions are then concatenated

to form the final feature set, ensuring that the multi-directional information is integrated into a comprehensive analysis.

$$[\mathbf{X}_{\text{Radial\_X, PCA}}, \mathbf{X}_{\text{Radial\_Y, PCA}}, \mathbf{X}_{\text{Axial, PCA}}]$$

PCA Example

Let's assume that we have the following data for Radial_X, Radial_Y, and Axial signals:

$$\mathbf{X}_{\text{Radial\_X}} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix},$$

$$\mathbf{X}_{\text{Radial\_Y}} = \begin{pmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \\ 8 & 9 & 10 \end{pmatrix},$$

$$\mathbf{X}_{\text{Axial}} = \begin{pmatrix} 3 & 4 & 5 \\ 6 & 7 & 8 \\ 9 & 10 & 11 \end{pmatrix}$$

Standardization

We standardize the features by subtracting the mean and dividing by the standard deviation using Equation (1). This step ensures that the data has zero mean and unit variance, which is necessary for PCA. The mean and standard deviation for $\mathbf{X}_{\text{Radial\_X}}$ are:

$$\mu_{\mathbf{X}_{\text{Radial\_X}}} = \begin{pmatrix} 4 & 5 & 6 \end{pmatrix},$$
$$\sigma_{\mathbf{X}_{\text{Radial\_X}}} = \begin{pmatrix} 2.449 & 2.449 & 2.449 \end{pmatrix}$$

Standardized $\mathbf{X}_{\text{Radial\_X, std}}$:

$$\mathbf{X}_{\text{Radial\_X, std}} = \frac{\mathbf{X}_{\text{Radial\_X}} - \mu_{\mathbf{X}_{\text{Radial\_X}}}}{\sigma_{\mathbf{X}_{\text{Radial\_X}}}} = \begin{pmatrix} -1.224 & -1.224 & -1.224 \\ 0 & 0 & 0 \\ 1.224 & 1.224 & 1.224 \end{pmatrix}$$

Covariance Matrix Calculation

We then compute the covariance matrix of the standardized features by using Equation (2). The covariance matrix helps in understanding the relationships between different features.

$$\Sigma_{\mathbf{X}_{\text{Radial\_X}}} = \frac{1}{n-1} \mathbf{X}_{\text{Radial\_X, std}}^{\top} \mathbf{X}_{\text{Radial\_X, std}} = \begin{pmatrix} 1.5 & 1.5 & 1.5 \\ 1.5 & 1.5 & 1.5 \\ 1.5 & 1.5 & 1.5 \end{pmatrix}$$

Eigenvalue and Eigenvector Computation

After that, we solve the eigenvalue equation to obtain eigenvalues and eigenvectors by using Equation (3). The eigenvalues indicate the variance explained by each principal component, and the eigenvectors provide the directions of these components. Assume that the eigenvalues are $\lambda_1 = 4.5$ and $\lambda_2 = \lambda_3 = 0$, with corresponding eigenvectors:

$$\mathbf{v}_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

with $\mathbf{v}_2$, $\mathbf{v}_3$ as any orthogonal vectors in the null space

Principal Component Selection

We then select the top k principal components in Equation 4 (let's choose $k = 1$). This step simplifies the data by reducing its dimensions while retaining the most significant variance.

$$\mathbf{V}_1 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}$$

Data Transformation

Finally, we transform the standardized Radial_X data into the principal component space using Equation (5). This transformation projects the data onto the selected principal components, resulting in a new representation of the data.

$$\mathbf{X}_{\text{Radial\_X, PCA}} = \mathbf{X}_{\text{Radial\_X, std}} \mathbf{V}_k = \begin{pmatrix} -1.224 & -1.224 & -1.224 \\ 0 & 0 & 0 \\ 1.224 & 1.224 & 1.224 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix} = \begin{pmatrix} -2.121 \\ 0 \\ 2.121 \end{pmatrix}$$

The resulting $\mathbf{X}_{\text{Radial\_X, PCA}}$ matrix represents the data projected onto the principal component, capturing the most significant variance in the data.

The cluster heatmap in Figure 8 derived from the Radial_X direction employs hierarchical clustering to group features with respect to the principal components, with the color-coding representing the degree of correlation. It corroborates the PCA's ability to concentrate variance and provides a visual interpretation tool to assess how different features inform the principal components. That results in showing that the multimodal feature extraction and the PCA dimension reduction offer a data-efficient, informative

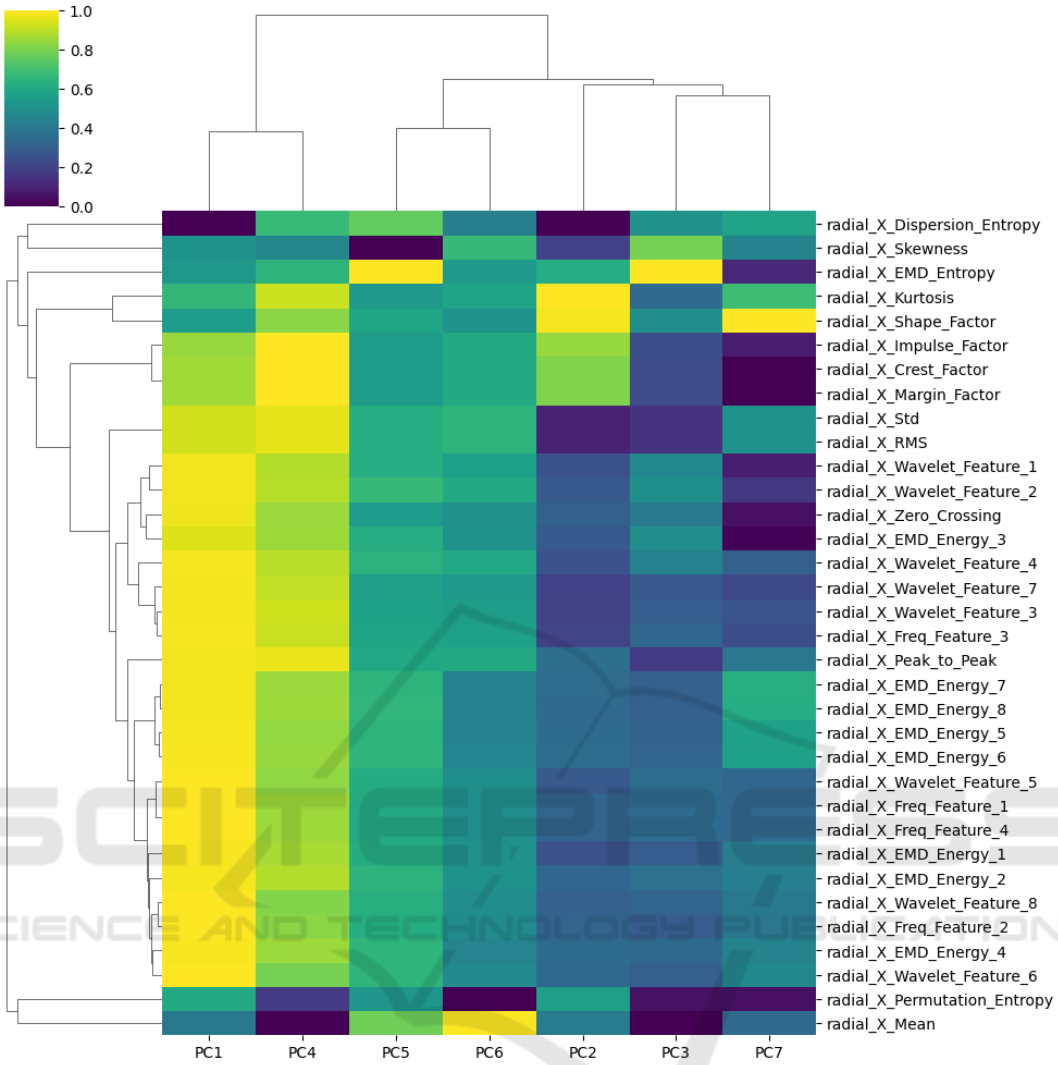

Figure 8: PCA Cluster Map.

foundation for transformer-based algorithms to detect and predict anomalies in pump systems. The visual representation through the cluster map further assists in interpreting the data, ensuring that the most significant features for anomaly detection are highlighted and utilized in predictive maintenance models.

When concatenating PCA features, we make the distinction between PCA Directions and PCA groups. The Directions approach involves applying PCA separately to features extracted from different measurement directions of vibration signals (Axial, Radial_X, and Radial_Y). Each direction's features result in distinct sets of principal components for each direction. This approach allows for a detailed examination of how different directions contribute to the overall condition or behavior being studied. On the other hand, the Feature Groups approach groups features based on their calculation methods or characteristics and then applies PCA to each group separately. The feature groups considered are statistical, EMD-Based, entropy measures, wavelet packet features, and frequency domain features. Each group's features are scaled and PCA is applied independently, resulting in distinct sets of principal components for each group. This method allows for a more nuanced understanding of the variability within each feature group and can highlight important patterns within those groups.

In both cases, the resulting datasets include metadata columns (Sensor_ID, Datetime, label) along with the PCA components for each group/direction.

4.3 Meta Transformer-Based Autoencoder

4.3.1 Transformer-AutoEncoder Anomaly Detection (TrAEAD)

In this model, shown in Figure 9, the input data in the form of a sequence of the PC vectors is used. The preprocessing and positional encoding are used to incorporate the positional relationship into the model and make the data in a format that is suitable for the model. This model aims to learn and enhance the discrimination of the representation of the data if the data is under the normal operation (denoted by D_t). After the training on the normal operation data is done, the reconstruction of this model should be the version of the normal operation of the input. If the difference between the input and the reconstruction input of the model is high, that implies that the input data is in a different distribution of the reconstruction data. In other words, the input data is anomalous.

This model achieves the task by using the concept of autoencoder where we aim to reconstruct any input. The green block is used for the encoding part, and the red-pink block is the fully-connected neural network blocks that are used for the reconstruction of the input. In the encoder block, the multi-head attention and fully-connected layers are used along with layer normalization and dropout. This is repeated N times for the scalability of the complexity of the model.

Mathematically, suppose we have a data input $x \in \mathbb{R}^p$ (or its isomorphic form like $x \in \mathbb{R}^{a\times b\times c} \cong \mathbb{R}^{abc}$) from the dataset $D = \{(x_1, y_1), \dots (x_n, y_n)\}, y_i \in \{0, 1\}$ that includes training data $D_t \subset \{(x, y) \in D | y = 0\}$ and validation data $D_v = D \setminus D_t$ and assume that latent space from the embedding function $f : \mathbb{R}^p \to \mathbb{R}^m$ is separable for each input x_i given the label y_i, that is, data from both classes are discriminative or distinct enough.

We can assume that the normal operation data is mostly deterministic with recognizable stochastic elements. Suppose the normal operation signal is $\phi(t)$. Then, the sensor, if perfect, should be able to fully capture the signal $\phi(t)$. However, as the sensor is not perfect, we can model this as $\phi(t) + \varepsilon(t)$ for measurement noise $\varepsilon(t) \sim T(\theta, t)$ for some distribution $T(\theta, t)$. However, the faulty operation will look different in either one of those as it will be $\phi(t) + \omega(t) + \varepsilon(t)$, which will have a different distribution for the faulty condition if $\omega(t)$ has enough impact. The assumption of discrimination can be used.

Now, suppose we create a model M that consists of two parts: an embedding function f and a reconstruction function $g : \mathbb{R}^m \to \mathbb{R}^p$. Given the training data $D_t \subset D$ where $D_t = \{(x_i, y_i) \mid y_i = 0\}$, the embedding function f maps this data to the latent space: $Z_t = \{f(x_i) \mid (x_i, y_i) \in D_t\} \subset \mathbb{R}^m$

Let $\hat{z} \in Z_t$ be a sample from the region in the latent space that is concentrated with normal operation data. The reconstruction function g then maps $\hat{z}$ back to the input space: $\hat{x} = g(\hat{z})$

Since $\hat{z}$ is from the image of f given the training data D_t, the reconstruction $\hat{x}$ should be similar to the original data x in D_t. Mathematically, this implies that:

$$\hat{x} = g(f(x_i)) \approx x_i \quad \text{for} \quad (x_i, y_i) \in D_t \tag{6}$$

Furthermore, we assume that normal operation data x outside of the training set D_t will be mapped by f to a neighborhood of Z_t in the latent space. Define the ε-neighborhood of Z_t as:

$$\mathcal{N}_\varepsilon(Z_t) = \{z \in \mathbb{R}^m \mid \exists z_t \in Z_t \text{ such that } \|z - z_t\| < \varepsilon\} \tag{7}$$

We assume that for normal operation data x outside of D_t:

$$f(x) \in \mathcal{N}_\varepsilon(Z_t) \quad \text{for some} \quad \varepsilon > 0 \tag{8}$$

This implies that for new normal operation data x, the embedding $f(x)$ lies within an ε-distance of the region Z_t in the latent space, ensuring that the reconstruction $g(f(x))$ is similar to normal operation data. This is possible due to our assumption that normal operation data is mostly deterministic with predictable stochastic elements. If the data is from the normal operation, the difference should be mostly on the deterministic part, and if f is good enough, the distance between the slight changes in value should be preserved.

Given the versatility and SOTA performance of transformer models in various domains (Vaswani et al., 2017; Raffel et al., 2020), it is reasonable to explore the possibility that f can be implemented as a transformer model. Transformers are capable of generating high-quality embeddings due to their self-attention mechanisms, which capture complex relationships and patterns in data (Vaswani et al., 2017).

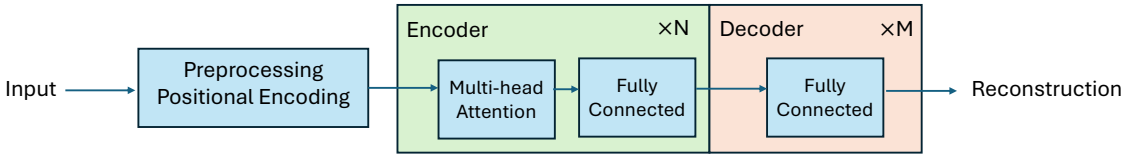

Figure 9: Architecture of TrAEAD Model.

However, it is important to note that while transformers have shown surprising empirical results, claiming that they will always preserve the distance for slight alterations of data points in the embedding space is an assumption that warrants further investigation. The empirical result later from (Wu et al., 2023) shows that the transformer model is a good fit for this task.

The transformer architecture is mainly drawn from (Wu et al., 2023). The basic fully connected layers are linked with multi-head attention to form a basic transformer model. While the authors use classification head on top of the transformer model to make it a classification task, in this work, the alteration has been done to make it a generative model to follow our previous assumption about the behavior of our input.

For the reconstruction part g, the primary requirement is that the reconstruction error with respect to the original data should be minimal. A simple neural network, such as a fully connected network, can be used with an objective function to minimize this reconstruction error. Given that our training data is a subset of the dataset containing only normal operation samples, the reconstruction error becomes a key factor in anomaly classification. Specifically, if the reconstruction error is high, it can imply that the data point is out of the training distribution and, thus, further from the neighborhood of the normal operation distribution. Therefore, it can be classified as an anomaly. In this study, the mean squared error (MSE):

$$\text{Error}_t = \frac{1}{|D_t|} \sum_{(x_i, y_i) \in D_t} ||x_i - M(x_i)||^2 \quad (9)$$

is used as the reconstruction error. The definition of "high" can be determined in the model selection phase, where metrics like Matthews' correlation coefficient (MCC) can be used as the cutting point. Specifically, the classification threshold that maximizes this metric would be the optimal threshold for classification. In other words, the threshold T is given by:

$$T = \arg\max_{T \in \mathbb{R}^+} \text{MCC}(\mathbb{I}(\text{Error}(x, M(x)) > T)) \quad (10)$$

where the indicator function $\mathbb{I}$ classifies whether the error between the true value x and the model's prediction $M(x)$ exceeds the threshold T.

4.3.2 Model Inference Example

Here is a simple example to illustrate the idea of how the model works. Suppose our input data is 3-axis accelerations from the accelerometer. For the sake of visualization, suppose we have a random vector $\mathbb{R}^3$ for three-dimensional signals with four samples. Here we have $x \in \mathbb{R}^{3\times 4}$ data.

Preprocessing

This data is then fed to the pipeline as 4 data points with three features each. That is for the signal

$$X = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

is treated as

$$x_1 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^\top$$
$$x_2 = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}^\top$$
$$x_3 = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}^\top$$
$$x_4 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\top$$

That is, we have three series of four or four lengths with three features. Then, for the fixed length size l, which is one of the hyperparameters of the preprocessing, the preprocessing will process a token as a group of vector $(x_1, \ldots, x_l), (x_{l+1}, \ldots, x_{2l}), \ldots$. Suppose that $l = 2$, then the first token for the model is (x_1, x_2), and the second token of the model is (x_3, x_4). It then will provide the positional encoding to each of the tokens. Specifically,

$$\text{PE}(i,j) = \begin{cases} \sin\left(\frac{i}{10000^{\frac{2}{d_{\text{model}}}\lfloor \frac{j}{2} \rfloor}}\right) & \text{if } j \text{ is even} \\ \cos\left(\frac{i}{10000^{\frac{2}{d_{\text{model}}}\lfloor \frac{j}{2} \rfloor + \frac{1}{d_{\text{model}}}}}\right) & \text{if } j \text{ is odd} \end{cases}$$

for the position i at dimension $j = 0, \ldots, d_{\text{model}}$. Here, $d_{\text{model}} = 3$. For the first token,

$$\text{PE}(1) = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^\top$$
$$\text{PE}(2) = \begin{bmatrix} 0.84147098 & 0.54030231 & 0.00215443 \end{bmatrix}^\top$$

and second token

$$\text{PE}(3) = \begin{bmatrix} 0.90929743 & -0.41614684 & 0.00430886 \end{bmatrix}^\top$$
$$\text{PE}(4) = \begin{bmatrix} 0.14112001 & -0.9899925 & 0.00646326 \end{bmatrix}^\top$$

The first input token to the transformer then is $(x_1 + \text{PE}(1), \ldots, x_l + \text{PE}(l))$.

Classification

After getting the reconstruction $M(X) = R_2(R_1(Z))$, given that the model is trained on the normal operation signal, the difference in input would be pronounced if the input and reconstruction are from different distributions. In this case, the MSE between $M(X)$ and X defined in Equation (9) is 0.38840. If the threshold from the training is larger than 0.38840, then this would mean X is under normal operation. However, if it is otherwise greater than the threshold, it is an abnormally.

Transformer Model (Encoder)

Suppose that $N = 1$; this input will then pass through the multi-head attention layer as outlined in (Vaswani et al., 2017). This attention is then combined with the input and passes through the fully connected layer. Note that the fully connected network works for each data point within the token. In other words, this fully connected layer works with the $\mathbb{R}^3$ vector for our example. For the simplicity of explanation, let the output for this model be the embedding in $\mathbb{R}^p$ and let $p = 5$. Suppose that the output for this encoder model is

$$Z = \begin{bmatrix} 0.144 & 0.213 & -0.115 & 0.821 & 1.110 \end{bmatrix}$$

Reconstruction (Decoder)

The series of fully connected networks in the decoder will map $\mathbb{R}^p \to \mathbb{R}^{nd_{\text{model}}}$ for n data points. In this case, it will map the embedding in $\mathbb{R}^p$ to $\mathbb{R}^{12}$ which is isomorphic with $\mathbb{R}^{3\times 4}$ which is the original data dimension. Suppose our $N = 2$ and we have $R_1 : \mathbb{R}^5 \to \mathbb{R}^8$ then $R_2 : \mathbb{R}^8 \to \mathbb{R}^{12}$ For the sake of space, the $\mathbb{R}^8$ will be written as $\mathbb{R}^{2\times 4}$ matrix and $\mathbb{R}^{12}$ will be written as $\mathbb{R}^{3\times 4}$.

The reconstruction will look like this

$$R_1(Z) = \begin{bmatrix} 0.811 & 0.911 & 0.314 & 0.112 \\ 0.000 & 0.213 & 0.981 & -0.772 \end{bmatrix}$$

$$R_2(R_1(Z)) = \begin{bmatrix} 0.811 & 0.591 & 0.994 & 0.991 \\ -0.012 & 0.818 & 0.011 & -0.772 \\ 0.012 & 0.995 & 0.892 & 0.742 \end{bmatrix}$$

5 EXPERIMENTAL STUDY

In these experiments, we are attempting to compare the performance of reconstruction anomaly detection from the (Wu et al., 2023) and (Yu et al., 2023) models to various data fusion methods paired with the TrAEAD model. These variations include testing the raw time wave-form data, the PCA Groups and PCA Directions methods, and the signal-averaged PCA Groups and PCA Directions approaches. For this comparison, we are focusing on metrics such as MCC score, Accuracy, Precision, Recall, F1 Score, Negative Predictive Value (NPV), and Specificity.

The assumption for the reconstruction approach is that the input data from the anomalous and non-anomalous sources should look different and significant enough to define a threshold of the reconstruction error to make a good classification decision. Using labeled data from the pumps at SCISTW, we are able to assess the performance of anomaly detection.

Since the accuracy of our models is based on the reconstruction error of the signals surpassing a predefined threshold to label a vibration reading as anomalous, the training set for this task only consists of data from the normally operating pumps (Pumps 5 and 6). This data was split into 80% training and 20% testing data, consisting of the first and last 10% by date recorded. The anomalous pump data from Pump 2 was reserved for testing.

The TrAEAD model reconstructs the signal well, with an overall reconstruction error from all the test signals of 0.01. However, the anomalous data has the tendency to be reconstructed with the wrong amplitude level. For the normal signal, the reconstruction fails to match some of the ground truth signals but manages to get the amplitude range correct.

As a baseline for comparison with the TrAEAD model, the data was reconstructed using a hybrid recurrent layer structure such as the model introduced by (Yu et al., 2023), originally intended for voiceprint anomaly detection, adapted to work with the time-waveform and spectrogram data, and the same reconstruction error threshold was used to label anomalies. Furthermore, the same experiment was repeated with the original model proposed by (Wu et al., 2023).

Table 1 shows that while all models perform well, the TrAEAD model with multimodal-feature extraction tends to perform better overall compared to the models by (Wu et al., 2023) and (Yu et al., 2023). Though these other models slightly outperform TrAEAD in Precision, NPV, and Specificity, the differences are marginal and don't significantly affect

Table 1: Performance Comparison of Various Models.

Data Fusion	Model	MCC	Acc.	Prec.	Rec.	F1	NPV	Spec.
Wavelet Scaleogram	Wu et al. (2023)	0.940	0.978	**1.000**	0.919	0.957	**0.971**	**1.000**
Raw Data	Yu et al. (2023)	0.921	0.962	**1.000**	0.943	0.971	0.898	**1.000**
Raw Data	TrAEAD	0.945	0.978	**1.000**	0.919	0.958	**0.971**	**1.000**
PCA Groups	TrAEAD	0.923	0.978	0.988	0.988	0.988	0.902	0.958
PCA Groups Sig. Avg.	TrAEAD	0.955	0.993	0.993	0.993	0.993	0.958	0.958
PCA Directions	TrAEAD	0.944	0.991	0.991	0.991	0.991	0.939	0.958
PCA Directions Sig. Avg.	**TrAEAD**	**0.966**	**0.995**	0.995	**0.995**	**0.995**	0.959	0.979

the model's overall value. TrAEAD's higher Recall and F1-score suggest it is better at identifying true anomalies and reducing false negatives, with the PCA Directions and Signal Averaging method producing the best results.

6 SEWAGE PUMP MONITORING SYSTEM PROTOTYPE

In order to host our META pipeline as well as visualize the output of analyzing pump vibration data, a web-based Sewage Pump Monitoring System was designed and built.

The system's backend was developed using Flask, a lightweight WSGI web application framework that provides tools for URL routing, request handling, and template rendering (Grinberg, 2018). For this application, Flask handles the upload and preprocessing of sensor data files, performs feature extraction on the raw sensor data, and exposes RESTful API endpoints that the frontend can interact with.

The frontend of the application, on the other hand, was built with React, a component-based JavaScript library for building user interfaces (Banks and Porcello, 2017). It integrates the backend APIs using asynchronous HTTP requests through libraries like Axios, and allows for the incorporation of visualization libraries such as Chart.js to display sensor data.

In this interface, the user is able to select and upload sensor data files with vibration data. This data is then processed through the META pipeline, from signal averaging, to multimodal feature extraction, and finally passed through the transformer-based autoencoder model. The output of this process is then displayed visually, organized by sensor, in a graph of amplitude versus time that shows three different lines for each dimension of the original data (Axial [pink], Radial_X [blue], and Radial_Y [green]) and marks the dates in which anomalies in the operation of the pumps are detected. The spans of consecutive data entries with anomalous data points are marked in a bright red color, which is toggle-able alongside the individual data from each of the input dimensions. A sample output for this process is shown in Figure 10.

7 CONCLUSION AND FUTURE WORK

In this paper, we propose a hybrid anomaly detection platform, called META, which integrates Multimodal-feature Extraction (ME) and a Transformer-based Autoencoder (TA) to leverage the strengths of the existing methodologies for predictive maintenance of sewage treatment plants. We first developed a signal averaging method that can preprocess the raw data using the signal processing algorithms to remove unrelated noise and improve the quality of signals related to the pump health and operations. Second, we enhanced the existing ME method to extract the meaningful signal properties on three vibration signal directions (Axial, Radial_X, and Radial_Y), respectively, by using five computational methods and measures and then fusing them together using PCA to generate PCA feature sets. Third, we enhanced a TA model to learn pump behavior from the extracted PCA feature sets to detect anomalous

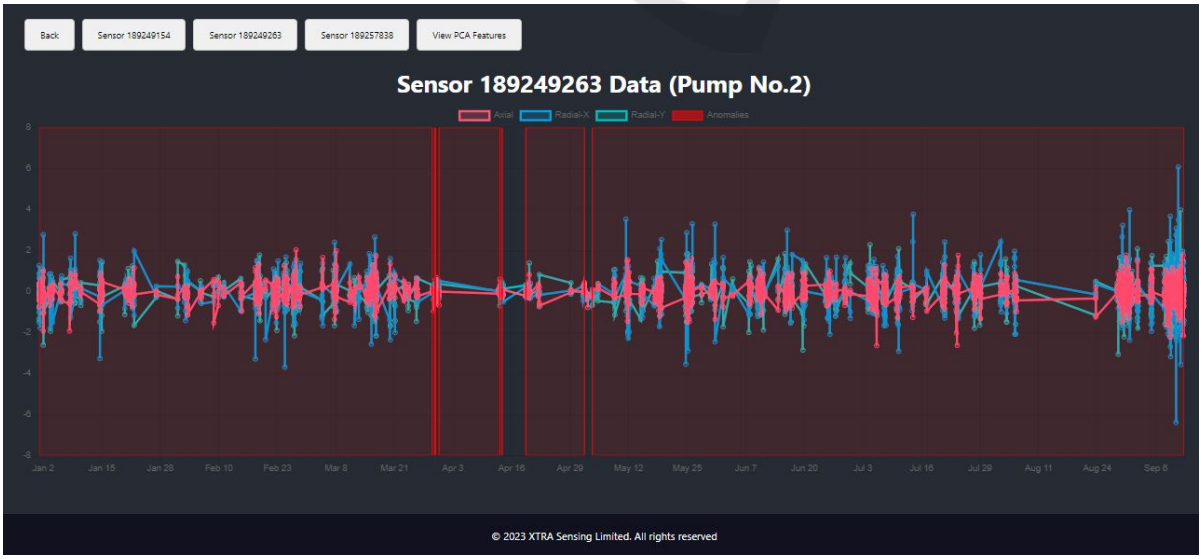

Figure 10: Sample GUI Output for the anomalous data, marked in red, from Pump No.2.

behavior over time with high precision. Fourth, we conducted an extensive experimental case study on the SCISTW, located in Hong Kong, to demonstrate that our META platform achieves SOTA performance in terms of MCC, F1-score, Accuracy, Precision, Recall, Negative Predictive Value, and Specificity. Finally, we built a prototype web-based Sewage Pump Monitoring System hosting the entire pipeline, providing an interactive user interface for future use.

Further research is needed to validate the versatility and robustness of the META framework. Assessing META's performance in anomaly detection for other types of industrial machinery as well as exploring different types of data fusion techniques could increase confidence in such a hybrid platform. For instance, examining late fusion techniques, in which the feature integration occurs at a later stage, right before the model makes a decision, could yield interesting insights into the system's performance. Furthermore, investigating the incorporation of other machine learning models with transformers and examining their impact on anomaly detection could lead to the development of more robust and scalable predictive maintenance approaches.

REFERENCES

Banks, A. and Porcello, E. (2017). *Learning React: Functional Web Development with React and Redux*. O'Reilly Media, Inc.

Carson, S. (2011). Best practice for lift stations: predictive maintenance or "run to fail"? Online: https://www.pumpsandsystems.com/best-practice-lift-stations-predictive-maintenance-or-run-fail.

Diez-Olivan, A., Ser, J. D., Galar, D., and Sierra, B. (2019). Data fusion and machine learning for industrial prognosis: Trends and perspectives towards industry 4.0. *Information Fusion*, 50:92–111. Full Length Article.

Drainage Services Department (2009a). Stonecutters Island Sewage Treatment Works. Online: https://www.dsd.gov.hk/TC/Files/publications_publicity/publicity_materials/leaflets_booklets_factsheets/Stonecutter.pdf. Last accessed on January 27, 2024.

Drainage Services Department (2009b). Stonecutters Island Sewage Treatment Works under Harbour Area Treatment Scheme Stage 2A. Online: https://www.dsd.gov.hk/EN/Files/publications_publicity/publicity_materials/leaflets_booklets_factsheets/HATS2A_Brochure_REV15.pdf. Last accessed on January 27, 2024.

Faisal, M., Muttaqi, K. M., Sutanto, D., Al-Shetwi, A. Q., Ker, P. J., and Hannan, M. (2023). Control technologies of wastewater treatment plants: The state-of-the-art, current challenges, and future directions. *Renewable and Sustainable Energy Reviews*, 181:113324. Available online 3 May 2023, Version of Record 3 May 2023.

Grinberg, M. (2018). *Flask Web Development: Developing Web Applications with Python*. O'Reilly Media, Inc.

Lahnsteiner, J. and Lempert, G. (2007). Water management in windhoek, namibia. *Water Science and Technology*, 55(1-2):441–448.

Lee, H. and Tan, T. P. (2016). Singapore's experience with reclaimed water: Newater. *International Journal of Water Resources Development*, 32(4):611–621.

Newton, E. (2021). Predictive maintenance for pump operations — modern pumping today.

Ormerod, K. J. and Silvia, L. (2017). Newspaper coverage of potable water recycling at orange county water district's groundwater replenishment system, 2000–2016. *Water*, 9(12):984. This article belongs to the Special Issue Development of Alternative Water Sources in the Urban Sector.

Pang, S., Yang, X., Zhang, X., and Lin, X. (2020). Fault diagnosis of rotating machinery with ensemble kernel extreme learning machine based on fused multi-domain features. *ISA Transactions*, 98:320–337.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(1).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. and Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA. Curran Associates Inc. [Google Scholar].

Wang, G., Zhao, Y., Zhang, J., and Ning, Y. (2021). A novel end-to-end feature selection and diagnosis method for rotating machinery. *Sensors*, 21(6).

Wu, H., Triebe, M. J., and Sutherland, J. W. (2023). A transformer-based approach for novel fault detection and fault classification/diagnosis in manufacturing: A rotary system application. *Journal of Manufacturing Systems*, 67:439–452.

Yu, D., Zhang, W., and Wang, H. (2023). Research on transformer voiceprint anomaly detection based on data-driven. *Energies*, 16(5).

Yuan, Z., Zhang, L., and Duan, L. (2018). A novel fusion diagnosis method for rotor system fault based on deep learning and multi-sourced heterogeneous monitoring data. *Measurement Science and Technology*, 29(11):115005.