HUMS2025 Data Challenge Result Summary

Team Name: Team A

Team Members: Dhiraj Neupane[1], Pradeep Kundu[2], Sunil Aryal[1], Richard Dazeley[1], Mohamed Reda Bouadjenek[1]

Institutions: 1: Deakin University, Australia 2: KU Leuven, Belgium

Publishable: Yes

1. Summary of Findings

For early convincing detection of a gearbox casing crack, the study utilised multi-channel vibration data, aiming to detect consistent patterns across channels. The earliest convincing detection of the fault was identified in the file ***'Day011_Hunting_SSA_Pinion_20240424_101221'***, using anomaly detection models (One-Class SVM, Isolation Forest, and a Convolutional Autoencoder). This study was inspired from Neupane et al. [1]. Further details on the methodologies, fault characteristics, and validation techniques are elaborated below.

2. Description of Analysis Methods

2.1 Preprocessing

The vibration dataset consisted of raw .mat files. Signals were normalized, segmented, and processed using the following steps:

1. **Signal Standardization**: Each signal was standardized (0 mean, 1 Std).
2. **Signal Averaging**: Signals were reshaped into P_SSA (684) and B_SSA (2556).
3. **Hilbert Transform and FFT**: Applied to extract envelope signals and isolate significant frequency components.
4. **Filtering**: Removed gear mesh harmonics and noise using custom harmonic removal and low-pass filtering.
5. **Residual Signal Extraction**: Obtained filtered signals representing fault-related information.
6. **Scalogram Image Generation:** After preprocessing, vibration signals were converted into scalograms using Continuous Wavelet Transform (CWT) with the Morlet wavelet.

2.2 Fault Detection Methods

Fault detection is vital for ensuring safe and uninterrupted industrial operations. This research employs semi-supervised anomaly detection (AD) methods, which rely on normal data for training, enabling the detection of unseen faults without requiring extensive labelled data.

Feature Extraction Using ResNet

ResNet50 was employed to extract features from scalogram images. ResNet50 was chosen due to its residual connections, which enable the efficient training of deep networks by mitigating the vanishing gradient problem. Scalogram images were processed through ResNet50, generating 2048-dimensional feature vectors from its second-last layer, capturing signal patterns.

Then the extracted feature vectors were then used as input for traditional AD models, including iF and OCSVM. These models were trained to identify deviations in feature distributions, with normal data serving as the baseline.

Training and Testing Data Partitioning

Training data from Days 3 to 7, assuming normal operations, was used to train the models, while testing data from Days 8 to 22, containing normal and abnormal signals, evaluated anomaly detection.

Semi-supervised approach (Training on Normal Data, testing on both)

1. ***One-Class SVM***:

OCSVM is a kernel-based method that constructs a hyperplane to separate normal data from anomalies in a high-dimensional space. ResNet50 was used to extract image features from scalogram representations of the processed signals. The model was trained on normal data, and anomalies were identified using decision thresholds (e.g., mean-3σ, 1st percentile and minimum value plus standard error).

2. ***Isolation Forest***:

Isolation forest is a tree-based algorithm that isolates anomalies by recursively partitioning the feature space. Data points requiring fewer splits to isolate are identified as anomalies. In this study, iF was applied to the ResNet50-extracted feature vectors. Anomalies were identified using thresholding techniques (e.g., mean-3σ, 1st percentile and minimum value plus standard error).

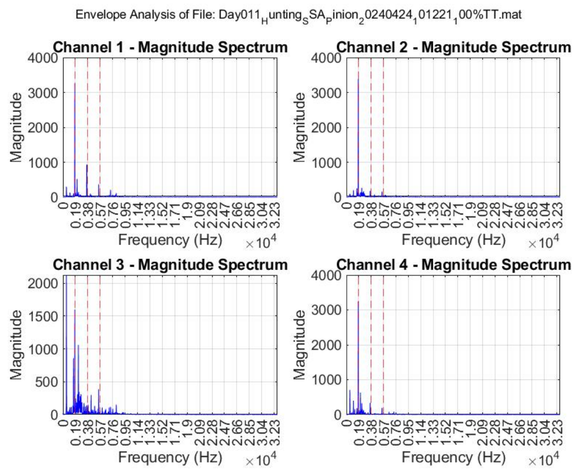
Unsupervised approach (Training on Normal Data, testing on both)

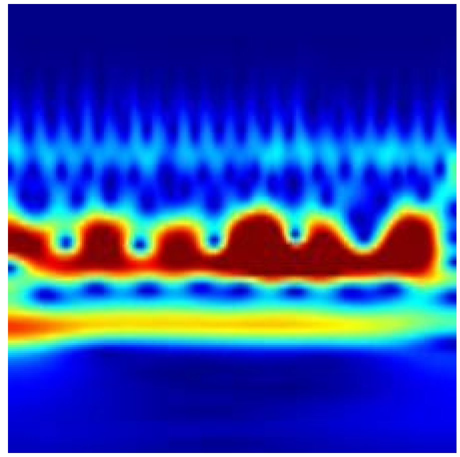
Convolutional Autoencoder:

The AE model was employed as an unsupervised learning technique, focusing on reconstructing the input data and identifying anomalies based on reconstruction errors. Unlike traditional methods, AEs do not rely on extracted features; instead, they operate directly on the input data (scalogram images). Reconstruction errors were calculated for all files, with anomalies defined as files exceeding the 99th percentile of errors. The AE model was trained on the entire dataset, learning to reconstruct normal scalogram images with minimal error.

3. Summary of Results

The integration of ResNet50 features with iF and OCSVM models identified the earliest convincing anomaly in file **'Day011 HuntingSSA_Pinion_20240424_101221'**. The use of AE provided additional insights into anomaly detection.





The above figures show the envelope analysis (Hilbert transform + FFT) & GMFs [left], and the scalogram image (CWT) [right] for the predicted file. Furthermore, since our approach relies on machine learning techniques—utilizing ResNet for feature extraction and Isolation Forest and One-Class SVM for anomaly detection—we have limited representative images available to illustrate the results. This is inherent to the methodology, which prioritizes feature analysis and anomaly prediction over generating visually interpretable outputs.

References

[1] Neupane, Dhiraj, et al. "A Comparative Study of Semi-Supervised Anomaly Detection Methods for Machine Fault Detection." *PHM Society European Conference*. Vol. 8. No. 1. 2024.