

*Keywords: predictive maintenance, audio spectrogram transformer, vibration analysis, deep learning*

Toukap Nono FERNAND JOSEPH <sup>1\*</sup>, Tokoue Ngatcha DIANORRÉ <sup>1</sup>,  
Offole FLORENC <sup>1</sup>, Mouzong Pemi MARCELIN <sup>1</sup>

<sup>1</sup> University of Douala, Cameroon, toukap\_nono@enspd-udo.cm, nonojoseph18@yahoo.com,  
florenceoffole33@gmail.com, mouz55@gmail.com

\* Corresponding author: toukap\_nono@enspd-udo.cm

## Modelling the predictive reliability of rotating machines using Artificial Intelligence

### Abstract

*In the context of Industry 5.0, predictive maintenance of rotating machinery is essential for improving operational efficiency, enhancing forecast reliability, and minimising unplanned downtime. This study presents an innovative approach that uses the Audio Spectrogram Transformer (AST) to predict the remaining useful life (RUL) of rotating machinery, leveraging a vibration signal dataset from a three-phase electric motor in the electrical engineering workshop at the IUT of Douala, called Vibration DATA IUT. By transforming vibration signals into time-series spectrograms, the AST model enables the detection of complex temporal patterns indicative of wear and degradation. Our methodology demonstrates superior predictive capabilities, achieving 98.9% accuracy and an impressive root mean square error (RMSE) of 1.2 hours in RUL estimation. These results highlight the effectiveness of integrating deep learning techniques with time-frequency domain analysis, offering an innovative framework for reliable failure prediction. These advances not only advance the state of the art in vibration analysis but also contribute significantly to predictive maintenance strategies in industrial applications, paving the way for increased safety and productivity.*

### 1. INTRODUCTION

The increasing complexity of industrial systems and the imperative of uninterrupted production have made predictive maintenance a fundamental strategy for ensuring the reliability, safety, and availability of industrial assets, particularly rotating machinery. In the context of Industry 4.0 and the emerging paradigm of Industry 5.0, industrial operators seek to maximise equipment performance while minimising maintenance costs and unplanned downtime (Priya et al., 2026; Hassan et al., 2024). Within this framework, the prediction of Remaining Useful Life (RUL) has become one of the most important prognostic tasks because it enables maintenance interventions to be planned before failures occur, thereby improving operational efficiency and reducing economic losses (Achouch et al., 2022).

Rotating machinery plays a crucial role in industrial production systems, and its degradation directly impacts productivity and system reliability. Vibration analysis is widely recognised as one of the most effective condition-monitoring techniques for assessing machine health and detecting degradation phenomena during operation (Burda et al., 2022). However, accurately estimating the remaining lifetime of equipment from vibration signals remains a challenging problem. Traditional vibration-analysis methods and prognostic approaches often rely on handcrafted features and expert knowledge, thereby limiting their ability to represent the complex, nonlinear degradation processes encountered in industrial environments (Zar et al., 2024). Moreover, rotating machinery frequently operates under varying loads and environmental conditions, leading to non-homogeneous data distributions and domain-shift problems that reduce the robustness and generalisation capability of conventional prognostic models (Garcia et al., 2025; Li et al., 2025). These limitations hinder the effective exploitation of real-time monitoring data for reliable RUL prediction and predictive maintenance decision-making.

Recent advances in artificial intelligence and deep learning have opened new opportunities for addressing these challenges. Advanced signal-processing techniques, super-resolution methods, and optimisation algorithms have demonstrated promising potential to improve industrial diagnostics and prognostics (Xu et al.;

Chardon, 2025). Likewise, meta-learning and transfer-learning strategies have been proposed to enhance model adaptability when operating conditions vary significantly or labelled degradation data are limited (Tuo, 2023). More recently, Transformer-based architectures have emerged as powerful tools for learning complex representations from sequential and time–frequency data due to their self-attention mechanisms (Li & Song, 2023; Tami et al., 2024). In parallel, spectrogram-based representations have proven effective for transforming vibration signals into informative time–frequency maps that facilitate the extraction of degradation-related features (Kong et al., 2020; Gong et al., 2021). These developments provide a promising foundation for the design of advanced prognostic models dedicated to Remaining Useful Life prediction.

Despite these advances, several important challenges remain insufficiently explored in vibration-based prognostics for rotating machinery. In particular, it is still necessary to determine whether Transformer-based architectures can effectively learn degradation patterns from vibration spectrograms to improve Remaining Useful Life prediction. Another open question concerns whether time–frequency representations can capture degradation information more effectively than conventional signal-processing approaches. Furthermore, the suitability of an adapted Audio Spectrogram Transformer for providing accurate and robust RUL estimates in predictive maintenance applications remains to be fully investigated. Based on these considerations, this study seeks to answer the following research questions: Can Transformer-based architectures effectively learn degradation patterns from vibration spectrograms to improve Remaining Useful Life prediction? Do time–frequency representations capture degradation-related information more effectively than conventional signal-processing approaches? Can an adapted Audio Spectrogram Transformer provide accurate and robust RUL estimates for predictive maintenance applications?

To address these challenges, this study is guided by the hypothesis that vibration spectrograms contain degradation signatures that can be effectively exploited by a Transformer-based architecture for Remaining Useful Life prediction. Accordingly, the objectives of this work are to develop a prognostic framework based on spectrogram representations, evaluate its ability to learn degradation-related information from vibration data, and assess its effectiveness in accurately estimating RUL in rotating machinery.

In this context, we propose a novel framework called NEW AST-VED (Audio Spectrogram Transformer for Vibration Event Detection), developed using vibration data collected in the Electrical Engineering Laboratory of the IUT and organised within the VIBRATION DATA IUT dataset. The proposed methodology transforms vibration signals into time–frequency spectrograms using the Short-Time Fourier Transform (STFT) and exploits an adapted Audio Spectrogram Transformer architecture to automatically learn degradation-related features and estimate the remaining useful life of rotating machinery.

The principal contribution of this work is therefore the development of a high-performance prognostic model for RUL prediction. Experimental results obtained on the VIBRATION DATA IUT dataset demonstrate the effectiveness of the proposed framework, achieving a Root Mean Square Error (RMSE) of 1.2 hours, a Mean Absolute Error (MAE) of 0.8 hours, and a coefficient of determination ( $R^2$ ) of 98.9%. These results indicate that the proposed approach successfully captures degradation dynamics and provides highly reliable prognostic information for predictive maintenance applications in Industry 5.0 environments.

To support the interpretation of the learned degradation representations, complementary fault-classification experiments were also conducted. The obtained accuracy of 98.9% and precision of 99.3% confirm the model's ability to distinguish among different degradation states and validate the relevance of the extracted features. However, these classification results serve only as supporting evidence of the quality of the learned representations, whereas the primary objective and main contribution of this study remain the accurate prediction of Remaining Useful Life (Tami et al., 2024).

The article is structured as follows: Section 2 reviews related work on reliability modelling and diagnostic methods, highlighting the innovation of our approach. Section 3 details the methodology used for RUL prediction, emphasising data management and the specific features of our model. Section 4 presents the results and discusses their implications for industrial predictive maintenance. Finally, Section 5 summarises the conclusions and outlines future research prospects.

## 2. RELATED WORK

Fault diagnosis and prognostics of rotating machinery have attracted significant research attention due to their importance in predictive maintenance and industrial reliability. Existing studies can generally be categorised into three major groups: convolutional neural network (CNN)-based methods, recurrent neural

network (RNN)-based approaches, particularly Long Short-Term Memory (LSTM) networks, Transformer-based architectures, and meta-learning approaches.

CNN-based methods have been widely adopted because of their ability to automatically extract informative features from complex vibration signals and time–frequency representations. For example, Zhang et al. (2025) developed a fault prediction method based on a fuzzy neural network in an Internet of Things (IoT) environment, demonstrating the effectiveness of deep learning techniques for industrial monitoring applications. Similarly, CNN architectures have been extensively applied to Remaining Useful Life (RUL) prediction using raw vibration signals or spectrogram-based representations (Zhao et al., 2021). Despite their effectiveness in feature extraction, CNN-based approaches may encounter difficulties in modelling long-range dependencies in degradation and adapting to highly dynamic industrial environments.

To overcome these limitations, recurrent neural network architectures have been extensively investigated for prognostic applications. In particular, Long Short-Term Memory (LSTM) networks have become a widely used baseline for RUL prediction due to their ability to capture temporal dependencies in degradation trajectories (Wu, Q et al., 2020; Huang et al, 2019; Wan et al., 2025). These models have demonstrated promising results in machinery health monitoring and prognostics. However, their performance may degrade when dealing with very long sequences or highly complex degradation patterns.

Recent advances in Transformer architectures have opened new perspectives for predictive maintenance and machine prognostics. Owing to their self-attention mechanisms, Transformer models can effectively learn relationships within sequential and time–frequency data without relying on recurrent processing. Zhou et al. (2023) proposed an explainable predictive maintenance strategy for multi-fault diagnosis in rotating machinery, highlighting the ability of Transformer-based models to capture complex relationships among degradation patterns. Similarly, several studies have successfully applied Transformer architectures to industrial signal analysis and vibration-based monitoring (Li & Song, 2023; Yousuf et al., 2022; Lei et al., 2025). These studies demonstrate the growing potential of self-attention mechanisms for extracting meaningful information from complex industrial data.

A particularly relevant contribution is the study by Huang et al. 2019, titled "Condition Monitoring in Industry 4.0 Production Systems: The Idea of Computational Intelligence Methods Application." Their work investigates the application of computational intelligence techniques for condition monitoring, anomaly detection, and maintenance support in Industry 4.0 production systems. While their study primarily focuses on fault detection and machine condition assessment, it does not directly address Remaining Useful Life prediction. In contrast, the present work focuses on prognostics, estimating the future degradation trajectory of rotating machinery via RUL prediction. Furthermore, our approach combines vibration spectrogram generation with an adapted Audio Spectrogram Transformer architecture specifically designed for regression tasks rather than fault classification.

In parallel, meta-learning approaches have emerged as promising solutions for improving the adaptability and generalisation capability of intelligent maintenance systems under varying operating conditions. Unlike conventional deep learning methods that require large amounts of labelled data for each new task, meta-learning aims to acquire transferable knowledge from multiple learning tasks, enabling rapid adaptation to previously unseen fault conditions with limited training data. (Tuo et al., 2023; Tian, et al., 2022) demonstrated that combining meta-learning and transfer-learning strategies can significantly enhance fault diagnosis performance in heterogeneous industrial environments. More broadly, recent state-of-the-art studies have shown that meta-learning techniques are particularly effective for addressing data scarcity, domain shift, and cross-domain fault diagnosis, which are common challenges in industrial applications (Toukap et al., 2025; Gharoun et al., 2024).

Several meta-learning frameworks, including few-shot learning and model-agnostic meta-learning approaches, have been successfully applied to machinery health monitoring, allowing models to generalise across different operating conditions and fault categories while reducing dependence on large labelled datasets (Vanschoren, 2019; Toukap et al., 2026). These approaches have shown promising results in fault detection, classification, and transfer diagnosis tasks by leveraging prior knowledge acquired from multiple degradation scenarios.

Despite these advantages, most existing meta-learning studies focus primarily on fault diagnosis and condition assessment rather than Remaining Useful Life (RUL) prediction. Furthermore, relatively few works have investigated the integration of meta-learning concepts with Transformer-based architectures and spectrogram-based vibration representations. Consequently, the potential of combining advanced feature-learning mechanisms with robust knowledge-transfer strategies for prognostic applications remains largely

unexplored. This limitation highlights an important research opportunity to develop more transferable and accurate RUL prediction frameworks that operate under diverse industrial conditions.

Therefore, despite substantial progress in fault diagnosis, condition monitoring, and predictive maintenance, the development of robust Transformer-based frameworks dedicated to vibration-based RUL prediction remains an open research challenge. Motivated by this gap, the proposed NEW AST-VED framework combines vibration spectrogram analysis with an adapted Audio Spectrogram Transformer architecture to model the evolution of degradation and accurately estimate the Remaining Useful Life of rotating machinery.

A relevant study closely related to our work is “Condition Monitoring in Industry 4.0 Production Systems the Idea of Computational Intelligence Methods Application” by Huang et al. 2019. Their work investigates the use of computational intelligence techniques for condition monitoring in Industry 4.0 environments, with a primary focus on fault detection, equipment condition assessment, and decision support for maintenance planning. The study highlights the potential of intelligent monitoring systems to identify abnormal operating conditions and improve maintenance efficiency.

In contrast, the objective of our work is not limited to fault detection or condition classification. We address the more challenging task of Remaining Useful Life (RUL) prediction, which requires estimating the time remaining before a machine fails. To achieve this objective, vibration signals are first transformed into time–frequency spectrograms using the Short-Time Fourier Transform (STFT), allowing the progressive degradation process to be represented visually and temporally. These spectrograms are then processed by the proposed NEW AST-VED architecture, which adapts the Audio Spectrogram Transformer (AST) to a regression framework specifically designed for RUL estimation.

Another important distinction lies in the learning strategy. While the work of Malmsten and Rantala focuses on computational intelligence methods for monitoring and fault identification, our approach leverages a transformer-based architecture trained directly on spectrogram representations associated with simulated RUL trajectories. Furthermore, the output layer of the AST model is modified from a classification-oriented structure to a dedicated regression layer, enabling continuous prediction of RUL values rather than discrete fault categories.

Therefore, although both studies contribute to intelligent condition monitoring within Industry 4.0 systems, the proposed NEW AST-VED framework extends beyond fault detection by integrating time–frequency signal transformation, transformer-based feature learning, and dedicated RUL regression capabilities for predictive maintenance applications.

### 3. EXPERIMENTAL DESIGN, MATERIALS AND METHODS

This section details the resources and techniques used to develop our model for predicting remaining useful life (RUL). Unlike fault classification approaches, our goal now is to predict a continuous value that represents the time remaining before a rotating machine reaches a critical failure state. We use a “transformer”-type model, NEW AST-VED, adapted for this regression task, based on the analysis of spectrograms derived from vibration data (VIBRATION DATA IUT). Figure 1 illustrates the general workflow of our predictive approach.

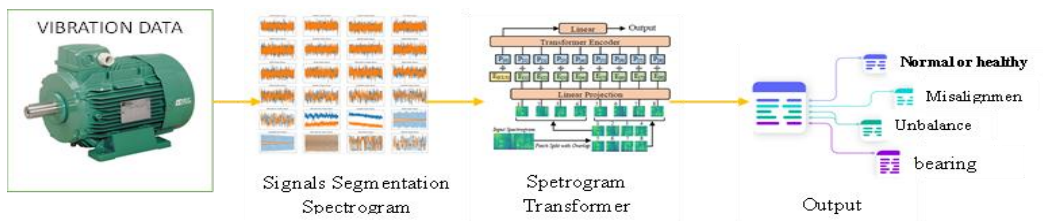


Fig. 1. Workflow of the proposed NEW AST-VED model for RUL prediction

#### 3.1. Description of the dataset and definition of the rule

Predicting remaining useful life (RUL) requires datasets that track machines' degradation trajectories until failure. Although the dataset is rich in fault classification, it is not inherently designed for RUL prediction, as it does not provide complete operational trajectories leading to failure and corresponding RUL values. For this

predictive study, we adopted an alternative approach to generate synthetic RUL data from the VIBRATION DATA IUT dataset.

The VIBRATION DATA IUT dataset contains 200,000 multivariate data points per class collected by sensors, covering five failure types (Normal operation, Ball defect, Ring defect, Bearing failure, Shaft defect) and a normal operating state. For each failure type, several severity levels are implicitly represented across the various “measurements” (Table 1).

**Tab.1. Specifications Table**

| Specifications              | Details                                                                                                        |
|-----------------------------|----------------------------------------------------------------------------------------------------------------|
| Data source                 | Three-phase electric motor in the electrical engineering workshop at the University Institute of Technology    |
| Sensor type                 | Piezoelectric sensor mounted on the motor shaft                                                                |
| Total number of data points | 1 600 000 raw data points                                                                                      |
| Sampling frequency          | 5000 Hz                                                                                                        |
| Defect classes studied      | 1. Normal operation                                                                                            |
|                             | 2. Ball defect                                                                                                 |
|                             | 3. Ring defect                                                                                                 |
|                             | 4. Bearing failure                                                                                             |
|                             | 5. Shaft defect                                                                                                |
| Levels of severity          | Speed 1, Speed 2 (except for normal operation)                                                                 |
| Points per class            | 200 000 points per class                                                                                       |
| Calculation of total data   | $5 \text{ faults} \times 2 \text{ levels} \times 200000 + 1 \text{ regular class} = 1.6 \text{ million lines}$ |
| Data acquisition            | Use of piezoelectric sensors to collect vibration signals                                                      |
| Pretreatment                | White noise, filtering, cleaning, windowing                                                                    |
| Data windows                | Fixed size N with or without overlap for each data segment                                                     |
| Filtering standards         | ISO 13373-1, VDI 3832: bandwidth from 10 Hz to 1000 Hz                                                         |
| Tools used                  | Python libraries (numpy, pandas, scipy)                                                                        |
| Expected results            | Identification of operating status and mechanical health                                                       |

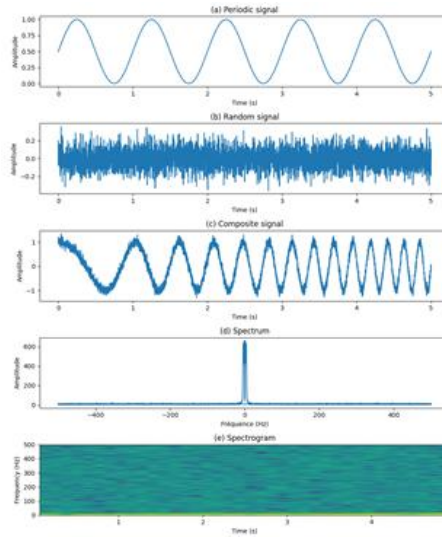
This study focuses on analysing vibration data acquired under various engine health and failure conditions, which constitute the operational variables of our investigation. All data was collected from a three-phase asynchronous electric motor (MAS) integrated into an experimental test bench, where vibration signals were measured using a high-sensitivity piezoelectric sensor (typical sensitivity of 100 mV for acceleration, with a useful bandwidth of 0.5 Hz to 10 kHz) strategically positioned on the motor shaft. An Arduino Uno microcontroller, equipped with a 10-bit analogue-to-digital converter (ADC), digitised these signals at a sampling frequency of 5 kHz. The overall dataset totals 1.6 million points, divided between nine distinct states: one state of normal machine operation and four types of specific mechanical faults (ball fault, ring fault, bearing fault, and shaft fault), each characterised by two distinct severity levels (‘speed 1’ and ‘speed 2’). Each condition contributed 200,000 data points, and the entire corpus was artificially enriched by adding Gaussian white noise to simulate more realistic industrial environmental conditions.

The vibration data, which is the main focus of this investigation, was collected using a highly sensitive piezoelectric sensor (characterised by a typical sensitivity of 100 mV/g and a useful bandwidth of 0.5 Hz to 10 kHz) strategically positioned on the shaft of a three-phase asynchronous electric motor (MAS), integrated into a dedicated experimental test bench. The analogue signal emitted by the sensor, with a maximum amplitude of 5 volts, was digitised by an Arduino Uno microcontroller equipped with a 10-bit analogue-to-digital converter (ADC). The digitisation produced raw values ranging from 0 to 1023, acquired at a sampling frequency of 5 kHz. This data was transmitted in real time to a laptop computer (Intel i5, 8 GB RAM, 3.5 GHz CPU) for subsequent preprocessing and analysis. The overall dataset totals 1.6 million points and is structured in a table comprising four distinct columns: ‘time’ (timestamp), ‘vibration’ (raw CAN values), ‘fault types’ and ‘severity level’. This corpus includes records for five distinct fault classes, including ball fault, ring fault, bearing fault, and shaft fault, each modelled with two distinct severity levels, in addition to a class representing normal machine operation. Each operating or fault state contributed 200,000 data points, and the entire dataset was artificially enriched by adding Gaussian white noise to simulate more realistic industrial environmental conditions. For the machine learning and model evaluation phases, this data was stored in separate CSV files (‘train.csv’, ‘val.csv’, ‘test.csv’), managed and manipulated using Python libraries (Numpy, Pandas, Scikit-

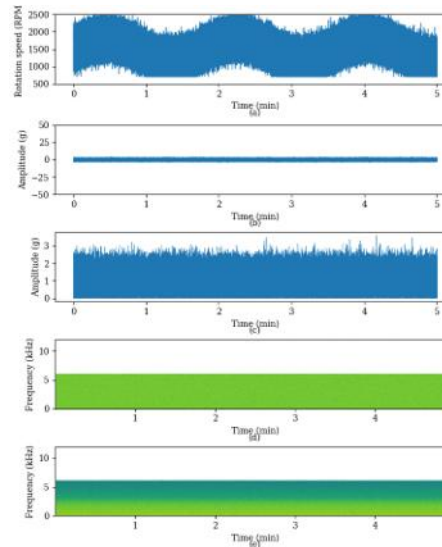
learn), ensuring a stratified and representative distribution of the different classes across the training, validation, and test sets.

The dataset collected for this study consists mainly of vibration measurements acquired from a three-phase asynchronous electric motor. This multivariate dataset includes recordings for five distinct classes of mechanical faults: normal operation, ball fault, ring fault, bearing fault, and shaft fault. Each type of fault, except normal operation, is characterised by two severity levels (“speed 1” and “speed 2”), enabling detailed analysis of failure progression. The vibration signals, captured by piezoelectric sensors, were sampled at 5 kHz and digitised, forming a corpus of 1.6 million raw data points, deliberately enriched with Gaussian white noise to increase the robustness of machine learning models in realistic industrial contexts.

The collected vibration and acoustic data are stored in binary Python files. The vibration data file contains five columns, namely Normal Operation, Ball Fault, Ring Fault, Bearing Fault, and Shaft Fault. The unit of vibration data is gravitational constant (g) ( $1g \approx 9.80665 \text{ m/s}^2$ ). To provide more detail in this dataset, synchronised velocity data is also included. Examples of raw data and their time-frequency analysis for each condition are shown in Figures 2 and 3.



**Fig. 2. Raw data and their time-frequency analysis in the x and y directions under normal conditions:**  
**(a) periodic motor rotation speed, (b) chronological speed data, (c) chronological speed spectrum,**  
**(d) spectrogram corresponding to speed, (e) spectrogram corresponding to speed**



**Fig. 3. Raw data and its time-frequency analysis in the x and y directions in a failure state: (a) periodic motor rotation speed,**  
**(b) chronological speed data, (c) chronological speed spectrum, (d) spectrogram corresponding to speed,**  
**(e) spectrogram corresponding to speed**

The hardware configuration and measurement setup are described in Section 3.1 and are not repeated here for brevity. On-site, vibration signals are collected using a piezoelectric sensor mounted on the motor shaft. This data collection system is essential for analysing motor performance and detecting anomalies.

### 3.2. Method of data generation and acquisition fault seedings

#### 3.2.1. Data acquisition test bench

The test bench is centred around a rotating machine, specifically, a three-phase electric motor installed in the IUT's electrical engineering workshop, as shown in Figure 4, on whose shaft a piezoelectric vibration sensor is mounted, as shown in Figure 5, an essential component for measuring vibration accelerations, as our study has demonstrated. An Arduino Uno microcontroller serves as an analogue-to-digital converter (ADC) (Figure 6); it converts the analogue signal from the sensor into a digital signal with an output voltage of 5 volts. The vibration data used in this study were acquired using the experimental setup described in Section 3.1.

Our entire test bench and data acquisition system are shown in Figure 7.



Fig. 4. Test bench

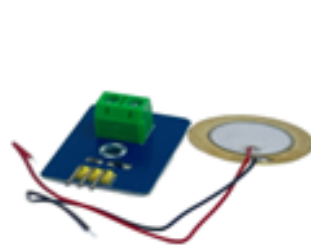


Fig. 5. Piezoelectric sensor

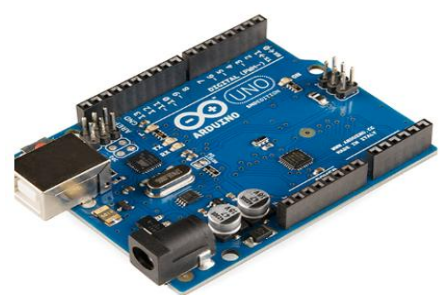


Fig. 6. Arduino Uno microcontroller

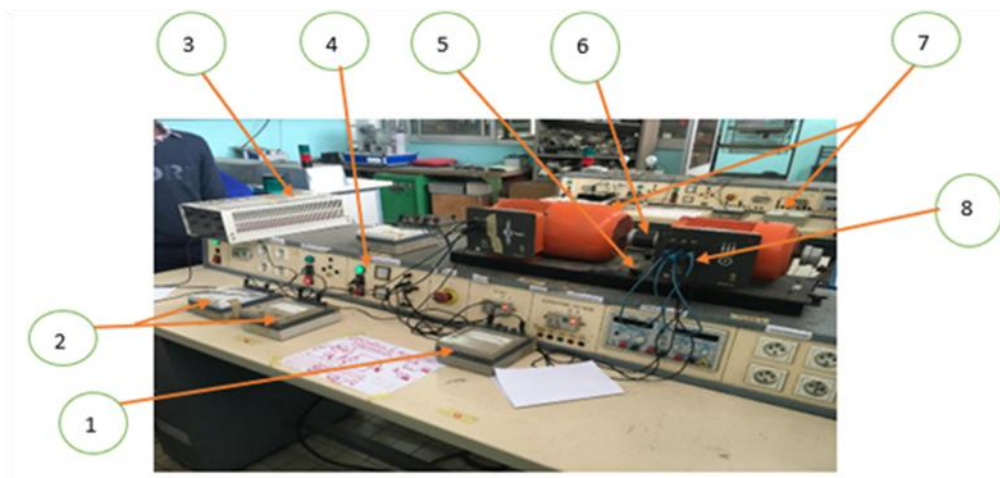


Fig. 7. Our entire test bench

- 1) Voltmeter
- 2) Ammeter
- 3) Rheostat
- 4) Power Source
- 5) Micro acoustic emission sensors
- 6) Rotor and shaft carrying the cooling fan
- 7) Rotating machine: Asynchronous Machine (MAS)
- 8) Terminal plate for coupling

### 3.2.2. Data acquired

The dataset includes raw vibration data collected under various engine operating conditions, with a total of 1,600,000 raw data points from a vibration signal  $s(t)$  along one axis, sampled at 5000 Hz. The dataset is divided into five main classes of mechanical faults, as shown in the Specie Table: normal engine operation, ball fault, ring fault, bearing fault, and shaft fault. Each class (except normal operation) is studied at two severity levels (“speed 1” and “speed 2”) to characterise the fault's evolution. For each fault condition and severity level, 200,000 data points were collected, and each data point is associated with white noise simulating realistic industrial conditions.

### 3.2.3. Data processing method

The thesis's methodology is a complete pipeline, from acquisition to modelling, integrating advanced signal processing and machine learning techniques.

- Modelling of the raw database: Creation of a data table with columns for time, vibration, fault type, and severity level, using the numpy and pandas libraries. The raw signal is modelled as a sum of sinusoidal components (main frequency, defect harmonics, random phases) and Gaussian white noise.
- Filtering and cleaning the vibration signal:
  - Bandpass filtering: Application of a bandpass filter (10 Hz–1000 Hz) in accordance with industry standards (ISO 13373-1, VDI 3832) to isolate the frequencies relevant to mechanical faults.
  - Silent/flat window suppression: Elimination of signal segments with low energy or variance to remove non-informative noise.
  - Signal centring: Removal of DC bias (offset) for a signal centred around zero.
- Window slicing: Division of the filtered signal into fixed-size segments ( $N=512$  points, corresponding to 100 ms at 5000 Hz), without overlap, to enable local analysis and statistical normalisation.
- Post-division cleaning: Application of criteria to retain informative windows (minimum standard deviation, minimum energy, reasonable peak/RMS ratio), eliminating spurious segments.
- Extraction of characteristic data: Calculation of relevant characteristics in the time and frequency domains for each window, as recommended by ISO 13373-2.

For each machine (identified by a fault type and severity level), we simulated a degradation trajectory. The RUL is defined as the time remaining before the machine's condition reaches a critical failure threshold, or, in our case, before the fault is fully established or a “faulty” state is reached. More specifically, for each vibration sample in a given defective state, we assigned a decreasing RUL from a maximum value (corresponding to the initial onset of the fault) to zero (corresponding to advanced failure). The “normal” state is associated with a maximum or infinite RUL, and fault states progress toward a zero RUL. This approach transforms a classification problem into a regression problem, where the model learns to correlate vibration characteristics with proximity to failure.

Each vibration sample, after conversion to a spectrogram, is thus associated with a numerical RUL value.

Adaptation of Tables 2 and 3

The original tables describe the distribution of defect classes. For RUL, we need to adapt this view to reflect the distribution of RUL values.

**Tab. 2. Distribution of defects and simulated RUL range in the dataset**

| Defects         | Simulated RUL range |
|-----------------|---------------------|
| Ball defect     | [0, 150]            |
| Ring defect     | [0, 240]            |
| Bearing failure | [0, 295]            |
| Shaft defect    | [0, 545]            |

**Tab. 3. Distribution of the dataset for training and testing (by machine ID/simulated trajectory)**

| Set type | Percentage | Number of simulated trajectories |
|----------|------------|----------------------------------|
| Training | 70 %       | X                                |
| Test     | 30 %       | Y                                |
| Total    | 100 %      | X + Y                            |

The training/testing split was implemented so that the degradation trajectories (or simulated machines) used for testing were distinct from those used for training, thereby avoiding data leakage and ensuring a realistic assessment of the model's generalisation ability.

### 3.2.4. Justification and limitations of the synthetic RUL generation process

The generation of synthetic Remaining Useful Life (RUL) trajectories was motivated by the absence of complete run-to-failure records in the VIBRATION DATA IUT dataset. Although the dataset provides rich information on machine health states, fault types, and severity levels, it does not contain continuous degradation histories associated with measured failure times. Consequently, a synthetic prognostic framework was established to enable supervised RUL learning.

The proposed degradation model is based on the widely accepted assumption that the deterioration of rotating machinery evolves progressively over time. As mechanical faults develop, vibration signatures generally exhibit increasing amplitudes, spectral distortions, harmonic components, and impulsive phenomena that reflect the loss of mechanical integrity. This monotonic degradation assumption is consistent with prognostic principles commonly adopted in predictive maintenance and reliability engineering, in which machine health gradually declines until it reaches a critical failure threshold.

For each fault category (ball defect, ring defect, bearing failure, and shaft defect), a degradation trajectory was simulated by assigning RUL values that decrease progressively from an initial maximum value toward zero. The maximum RUL values reported in Table 2 (150 h, 240 h, 295 h, and 545 h) were selected based on conservative engineering estimates that reflect the relative severity and expected progression of degradation for the corresponding fault mechanisms.

The validity of this approach is supported by the structure of the VIBRATION DATA IUT dataset itself. Each fault category is represented across multiple severity levels and operating conditions, enabling the construction of degradation sequences that preserve the natural ordering among healthy, intermediate, and severely degraded machine states. Rather than assigning arbitrary labels, the synthetic RUL values were designed to follow the progression of fault severity observed in the vibration measurements, ensuring consistency between signal degradation and remaining lifetime estimation.

To further assess the realism of the generated trajectories, their characteristics were compared with degradation behaviours commonly reported in the predictive maintenance literature. The resulting trajectories exhibit a gradual deterioration phase followed by accelerated degradation as the machine approaches the failure threshold, which is consistent with degradation patterns observed in bearings, shafts, and rotating electrical machines. Similar assumptions have been widely adopted in prognostic studies when real run-to-failure data are unavailable, particularly for the development and benchmarking of data-driven RUL prediction models.

Nevertheless, it is important to acknowledge that the generated RUL labels remain synthetic and cannot fully reproduce the variability of real industrial degradation processes. Therefore, the reported predictive performance (RMSE = 1.2 h, MAE = 0.8 h,  $R^2 = 98.9\%$ ) should be interpreted within the context of the proposed simulation framework. Future work will focus on validating the NEW AST-VED model using real run-to-failure datasets to confirm its generalisation capability under practical industrial conditions.

## 3.3. Data pre-processing

Pre-processing steps are crucial to ensure the quality and relevance of vibration data for RUL prediction. The Vibration dataset IUT has already undergone pre-processing for noise reduction. We supplemented these steps with additional methods to optimise the data for our prognostic task.

### 3.3.1. Additional pre-treatment

The raw data, loaded in CSV format using Python's NumPy tools, was visualised to identify irregularities. The labelling of the data with the appropriate defect states was verified. To improve signal quality, we applied the following techniques:

- Removal of outliers: Identification and elimination of extreme data points that could skew the degradation analysis.
- Noise filtering: Use of a notch filter to remove unwanted noise frequencies, ensuring that signal variations are primarily due to machine dynamics.

- Smoothing of health state indicators (HSI): If health indicators are derived from vibration signals, smoothing (e.g., moving average or Kalman filter) is applied to obtain more stable and less noisy degradation trajectories, facilitating RUL learning.

### 3.3.2. Normalization

To harmonise the scales of the different vibration characteristics and ensure stable convergence of the model, min-max normalisation was applied to the entire data set. This method adjusts the values of the vibration signals within a defined range, generally between 0 and 1, using the following equation 1:

$$X_n = \frac{X - \text{Min}(X)}{\text{Max}(X) - \text{Min}(X)} \quad (1)$$

Where:  $X$  – normalise  
 $X_n$  – is a vector in the form of an array.  
 $\text{Min}(X)$  – This is the lowest value of the vector  $X$ .  
 $\text{Max}(X)$  – This is the highest value of the vector  $X$ .

### 3.4. Feature extraction

The feature extraction phase is essential for transforming the raw vibration data (VIBRATION DATA IUT) into a representation suitable for the new AST-VED model. This step aims to prepare the data for RUL regression.

The first step involves transforming the raw time-domain data into the frequency and time-frequency domains using Fourier transforms. The discrete Fourier transform (DFT) is applied to analyse the frequency composition of the segmented signals. Equations (2) and (3) describe the DFT and its inverse transform:

$$X [K] = \sum_{n=0}^{N-1} x[n] e^{-j \frac{2\pi}{N} kn} \quad (2)$$

$$x [k] = \frac{1}{N} \sum_{n=0}^{N-1} X[n] e^{-j \frac{2\pi}{N} kn} \quad (3)$$

Where:  $X [K]$  – the coefficient of the DFT at index  $k$   
 $x[n]$  – the input signal at index  $n$   
 $N$  – the total number of signal samples.  
 $j$  – the imaginary unit.  
 $k$  – Varies from 0 to  $N - 1$

These transformations enable the construction of features in the time, frequency, and time-frequency domains, which are then used to generate spectrograms.

#### 3.4.1. Selection of characteristics

Although conversion to spectrograms already reduces the need for manual feature selection, a correlation analysis can be performed to identify and potentially remove redundant or irrelevant features if additional features are calculated. The SelectKBest algorithm can be used to identify the most relevant features. Dimension reduction, such as Principal Component Analysis (PCA), can also be applied to compress information and reduce computational complexity.

### 3.5. Data transformation into spectrograms

The conversion of vibration data into spectrograms is a pivotal step in our methodology, as it allows vibration signals to be analysed in the time-frequency domain, a rich representation for RUL prognosis. For

effective application of the Audio Spectrogram Transformer (AST) model, four fundamental modifications are made to the pre-processed vibration data (Fang et al., 2023).

The first modification is resampling the data to 16 kHz. This choice is an essential compromise between computational efficiency and the preservation of vibration data quality (Xiao et al., 2025). This resampling frequency ensures that the vibration data retains the information required by the model while optimising processing power.

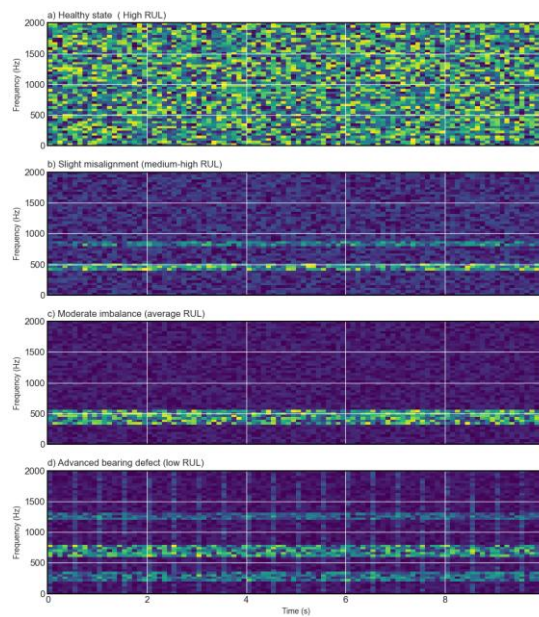
The second step is to apply the Short-Time Fourier Transform (STFT) to the resampled vibration signals. The STFT decomposes the signal into overlapping temporal windows and computes the frequency spectrum within each window, producing a time–frequency representation that captures non-stationary vibration characteristics. The resulting spectrograms preserve both temporal and spectral information, making them particularly suitable for monitoring the evolution of machine degradation.

The generated spectrograms provide a compact representation of vibration dynamics and reveal frequency patterns, transient events, and degradation-related structures that are often difficult to identify directly in the time domain. Such representations are especially advantageous for Transformer-based architectures because they enable the self-attention mechanism to learn relationships between temporal and spectral features simultaneously. Consequently, spectrogram generation constitutes a critical step in the proposed AST-VED framework for Remaining Useful Life (RUL) prediction.

The fundamental advantage of converting vibration data into spectrograms is that it provides an advanced visual representation that allows simultaneous access to both time- and frequency-domain information about events. This greatly facilitates the detection of patterns and anomalies that are typically indicative of problems with rotating gear, but which would be difficult to spot in the time domain alone. For RUL prognosis, these spectrograms capture the evolution of the vibration signature as the machine deteriorates, which is crucial for modelling the progression of failure.

This analytical approach has several advantages. This transformation offers several advantages for prognostic modelling. First, it enhances the representation of degradation-related information by jointly preserving temporal and frequency characteristics of vibration signals. Second, by converting vibration measurements into structured time–frequency maps, it enables the exploitation of advanced deep-learning architectures originally developed for image and audio analysis. This facilitates automatic feature extraction and improves the model's ability to identify degradation patterns associated with the remaining useful life of rotating machinery (Tami et al., 2024; Zhang et al., 2025).

The use of spectrograms makes analysis more reliable and efficient, particularly by bridging the gap between the robust and well-established field of image analysis and the dynamic and complicated nature of signals, as shown in Figure 8. For RUL prediction, spectrograms serve as the input “images” that the transformer analyses to infer the degradation state.



**Fig. 8. Vibration data spectra illustrating degradation states for RUL prediction**

The generated spectrograms provide a compact time–frequency representation of vibration signals, enabling the visualisation of frequency variations associated with machine degradation. Unlike raw time-domain signals, spectrograms reveal transient phenomena, harmonic components, and frequency patterns that are often indicative of developing faults in rotating machinery. This representation is particularly well-suited to Transformer-based architectures, as it preserves both temporal and spectral information while enabling automatic feature extraction via self-attention mechanisms. Consequently, the spectrogram transformation constitutes a critical step in the proposed AST-VED framework for vibration-based Remaining Useful Life (RUL) prediction.

### 3.6. Architecture of the transformer spectrum

The Audio Spectrogram Transformer (AST) architecture is at the heart of our method for analysing vibration data from rotating machines, drawing inspiration from the Vision Transformer (ViT) (M. Tami et al., 2024). For RUL prediction, the basic AST architecture is retained, but the output layer is specifically adapted for regression rather than classification. Our model, called NEW AST-VED, leverages the transformers' ability to process spectrogram sequences to model the progression of degradation and estimate RUL.

In the ViT model, images are divided into patches, and each patch is linearly projected before positional encodings are applied, thereby preserving the spatial relationships within the image. A sequence of vectors is then fed into a transformer encoder, which uses self-attention to capture complex interactions among different image segments. The encoder produces a higher-dimensional representation of the original image by capturing local and global dependencies. This approach is directly transferred to our model for vibration data.

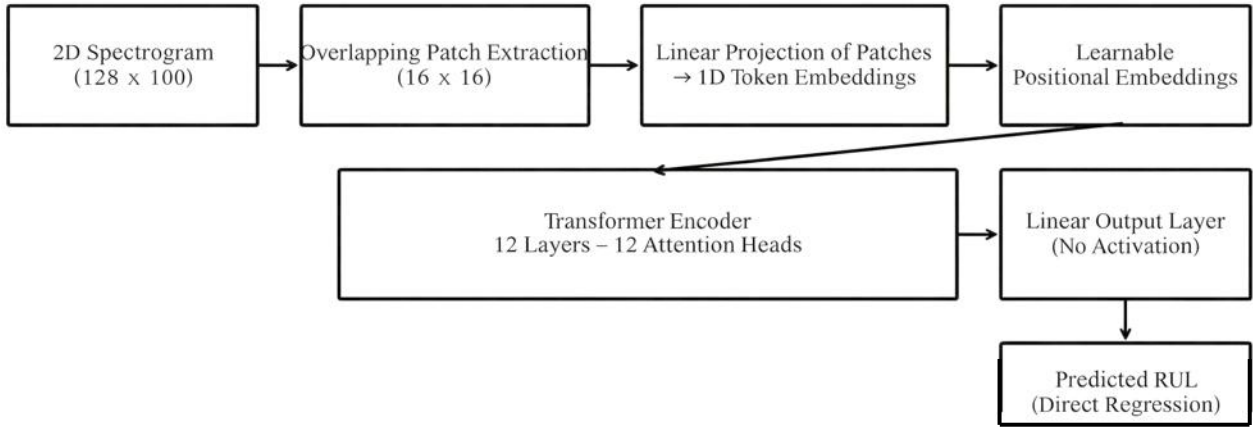
Compared to conventional machine learning techniques, transformers offer several advantages for RUL prediction. They are particularly effective at processing time-series data or spectrograms, as they can handle sequential data without requiring the explicit extraction of temporal information. Their self-attention mechanism allows them to effectively capture long-term dependencies in the input sequence, which is vital for modelling the progression of deterioration and predicting RUL. This capability surpasses that of recurrent neural networks (RNNs) and long short-term memory (LSTM) networks in both training time and in capturing spatial correlations in spectrograms (Li & Song, 2023).

The AST architecture enables the diagnosis of rotating machines by converting vibration spectrograms into an enhanced representation. By using positional embeddings, the power of the Transformer encoder, and dividing spectrograms into blocks, the model effectively captures critical connections in space and time for anomaly identification. For RUL, this capability is extended to detect degradation patterns.

The initial AST model divides a signal into 16x16 pixel patches and converts it into a 128x100 spectrogram. These patches are then connected to position embeddings after being transformed into vibration spectrograms (2D images). Spectrograms are divided into small parts called patches. To preserve spatial information, patch embeddings are associated with position embeddings. Each patch is flattened and then projected onto a smaller location. We use multiple transformer layers to capture relationships among patches. A feedforward neural network with multi-head attention is incorporated into the transformer at each level. A standard 12-layer, 12-head transformer encoder, designed for classification, processes the completed sequence.

Crucial change for RUL: The transformer's output layer is modified. Instead of a linear layer with sigmoid activation for classification, we use a linear layer without activation for direct RUL regression. This layer takes the aggregated representation of the spectrogram sequence and maps it to a single numerical value representing the predicted RUL.

This technique exploits the wealth of information contained in vibration spectrograms, which capture the frequency and temporal properties of signals. The Transformer's architecture facilitates the automatic extraction of salient features and the accurate identification of anomalies by capturing complex interactions among different parts of the spectrogram. Its adaptability and flexibility make it a flexible option for the automated diagnosis of rotating machines, paving the way for effective predictive maintenance and, in our case, accurate RUL prediction. The internal composition of the Audio Spectrogram Transformer (AST), also known as NEW AST-VED for detecting vibrational events, is shown in Figure 10 and is detailed below.



**Fig. 9. Proposed architecture of the Audio Spectrogram Transformer for RUL prediction (NEW AST-VED).** NEW AST-VED stands out for its robustness and efficiency in industrial environments, offering accurate detection of anomalies and applications that are crucial in terms of safety and reliability

Key architectural features for RUL:

- Spectrogram Size: The dimensions of the spectrogram (128 x 100) allow for detailed visualisation of temporal and frequency characteristics, crucial for tracking degradation.
- Patch Size: Each 16 x 16 pixel patch, with a 50% overlap, maximises feature extraction while preserving spatial relationships within the degradation signature.
- Transformer: Composed of 12 layers and 12 attention heads, it allows for in-depth modelling of complex relationships in vibratory data, including long-term dependencies necessary for RUL prediction.
- Regression Layer: The final linear layer, without activation, produces a continuous numerical value, the predicted RUL.

### 3.6.1. Representation of inputs by a patch

The method of representing inputs using patches, specifically for spectrograms, involves dividing the input into distinct areas called “patches.” The input spectrogram is separated into smaller tiles or squares that may overlap. Each patch is then transformed into a two-dimensional vector by rearranging the pixel values. Each flattened patch vector is then transposed into a vector space of dimensions masked by a linear layer. The next step is position integration.

To maintain spatial links between patch integrations, location data is incorporated into them. This is necessary because the placement and layout of patches can contain important information. The resulting sequence of vectors, each representing a patch in the original input, is then used as input data by the model, which functions similarly to a transformer. This approach allows the model to integrate the local and global links in the data, which are essential for understanding the progression of degradation.

### 3.6.2. Transformer encoder

The transformer encoder, a key component of transformer-based structures, is fundamental to our approach. Its main purpose is to convert an input sequence (in this case, a sequence of spectrogram patches) into a vector representation that encapsulates the relationships and context between its constituent parts. To account for the sequence, we first transform the patches into “embeddings,” then add location data. The encoder uses feedforward neural networks to improve the representations after an auto-attention system has evaluated the importance of each element relative to the others. The use of layer normalisation and residual connections improves training stability. This approach allows the encoder to process patches sequentially, unlike traditional convolutional neural networks that analyse spectrograms in their entirety. To identify distinct anomalies that indicate different moving machine problems and, more importantly, to model the progression of degradation for RUL prediction, the model must understand the temporal dynamics of the data, which is made possible by this sequential approach.

### 3.6.3. Linear regression layer

In our RUL prediction model architecture, the final linear layer is crucial. It takes the high-level representation generated by the transformer encoder and maps it directly to a continuous numerical value.

The final model predicts the RUL value. Unlike classification, where a sigmoid or softmax activation is used to obtain probabilities, RUL regression uses a linear layer without an activation function (or with an implicit linear activation). This allows the model to produce a continuous output that can represent any RUL value.

This result represents the RUL estimate for the machine state corresponding to the input spectrogram. The transformer's ability to recognise complex patterns in spectrogram data related to the progression of degradation improves RUL prediction accuracy compared to traditional techniques.

### 3.7. Training and evaluation measures of the model

For RUL prediction, the model's evaluation and training criteria are adapted to the regression task.

Optimisers such as SGDM are used for training. The loss function is chosen to minimise the error between the predicted RUL and the actual RUL.

#### 3.7.1. Loss function and evaluation criteria

Equations 4 are used for loss functions suitable for the regression of our RUL:

Mean Squared Error (MSE): a commonly used loss function for regression that penalises large errors more heavily.

$$L_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (4)$$

For our study, we primarily used MSE for training, as it encourages the model to make more accurate predictions and avoid significant errors.

To evaluate the regression performance of our study, we used metrics highly relevant to RUL prediction, as illustrated in equations 5, 6, 7, and 8.

Mean Absolute Error (MAE): Measures the average absolute difference between predictions and actual values. A lower MAE indicates better performance.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (5)$$

Root Mean Squared Error (RMSE): The square root of the MSE. It is sensitive to large errors and uses the same units as the predicted variable, making it more interpretable.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (6)$$

RMSE (Root Mean Square Error): Measures the mean square error, indicating how much the values predicted by the model deviate from the observed values.

Coefficient of Determination ( $R^2$ ): Indicates the proportion of the variance in the actual RUL that can be predicted from the input variables. An  $R^2$  close to 1 indicates a model that explains the variance in the data well.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (7)$$

Where  $R^2$  is the average of the actual RUL values. It measures the difference between the observed values  $y_i$  and the predicted values  $\hat{y}_i$ .

- $y_i$ : Observed value.
- $\hat{y}_i$ : Value predicted by the model.

**Prediction Score:** To evaluate performance in a predictive maintenance context, we can use scores that penalise early and late predictions differently, such as those proposed in prediction and health management competitions (PHM Society). A common score is defined as follows:

$$\text{Score} = \sum_{i=1}^N S_i \quad (8)$$

Where equation xx gives us the score for each prediction:

- $N$ : Total number of items.
- $S_i$ : Individual value for a score.

**Prediction Horizon:** The model's ability to accurately predict RUL at different time horizons is also evaluated. This involves analysing the model's performance when the actual RUL is high (indicating early degradation) and when it is low (indicating near-failure).

To determine the predicted reliability status of our predictive model, equations 9 and 10 provide the reliability function  $R(t)$ , which indicates how a machine's reliability changes over time, and the risk function, which shows how the risk of failure varies over time.

$$R(t) = 1 - F(t) \quad (9)$$

$R(t)$ : the survival function at time. It indicates the probability that an event (such as a failure) has not occurred up to that time.

- $F(t)$ : the distribution function or cumulative distribution function. It indicates the probability that the event has occurred up to time  $t$ .

$$\lambda(t) = \frac{f(t)}{R(t)} \quad (10)$$

- $\lambda(t)$ : the risk rate or failure rate at time  $t$ . This measures the instantaneous probability that an event (a failure) will occur at that given moment.
- $f(t)$ : the probability density function of the event at time  $t$ . This describes the probability of the event occurring at that precise moment.

The concept of class imbalance is less relevant for direct RUL regression. However, an imbalance in the degradation phases (lots of healthy data, little data near failure) can be challenging. To remedy this, sampling or loss-weighting techniques can be applied to give greater weight to samples that represent critical degradation phases.

The model was trained on an HP laptop equipped with an A100 GPU, allowing for rapid optimisation. The Transformers library was used to implement the model. The data distribution for training and testing is 70% and 30%, respectively, with separation by simulated trajectory ID to ensure independence between sets.

**Model Complexity and Computational Efficiency:** The model is based on the Transformer architecture and inherently complex, as it can capture long-term dependencies in sequential data. Even with this complexity, the model converges quickly, reaching optimal performance in just 10 epochs thanks to an adjusted learning rate and L2 regularisation. Training time on an A100 GPU was approximately ten minutes. Computational efficiency was improved through parallel data processing and rebalancing techniques, ensuring robust performance even in the presence of degradation-phase imbalance.

### 3.8.2. Measuring processing time

To evaluate the operational efficiency of our approach, we measured the time required for each phase of data processing:

- Data pre-processing: 30 seconds.
- Transformation into spectrograms: 25 seconds.
- Inference time (prediction of RUL on test data): 11 seconds.

These measurements highlight the speed of our method, which is essential for real-time industrial applications.

## 4. RESULTS, DICUSSIONS AND COMPARISON

This section presents and analyzes the results obtained by our model for predicting the Remaining Useful Life (RUL) of rotating machines. We will evaluate the model's performance using appropriate regression metrics and discuss its implications for industrial predictive maintenance, comparing it to existing approaches and highlighting its robustness.

The experiments were conducted on a workstation equipped with an Intel(R) Core i7-3320M 3 GHz processor and an NVIDIA A100 graphics card with 11 GB of buffer memory, using Python for program development and execution. In-depth hyperparameter optimization was performed to ensure optimal performance and stable convergence of the model for the RUL regression task.

### 4.1. Results

It should be emphasized that all experiments reported in this study were conducted exclusively using the VIBRATION DATA IUT dataset developed at the IUT Electrical Engineering Laboratory. The fault categories, degradation levels, vibration measurements, and synthetic RUL trajectories were all derived from this dataset. No samples from external datasets were used for model training, validation, or testing. Any references to benchmark datasets in the manuscript are provided solely for methodological comparison and do not constitute part of the experimental data used in this work.

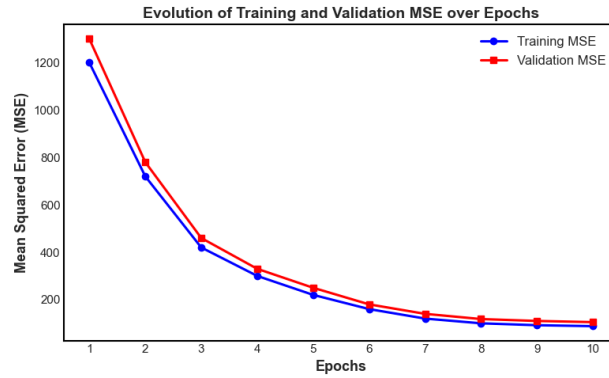
#### 4.1.1. Hyperparameter optimization and training progression

Identifying the optimal hyperparameters is crucial for the performance of the RUL prediction model. The hyperparameters were tuned using a predefined search space covering the learning rate ( $1 \times 10^{-5}$  to  $1 \times 10^{-3}$ ), weight decay (0.1% to 1%), batch size (8 to 64), optimiser (SGDM, Adam, RMSprop), and learning-rate scheduler (linear, cosine). This space was first explored with a Randomised Search over 50 configurations to identify promising regions, followed by a refinement stage using Bayesian Optimisation (Hyperopt, Tree-structured Parzen Estimator) over 30 additional trials, using the validation RMSE as the objective function. Following this procedure, we selected the configuration shown in Table 4. A learning rate of  $7 \times 10^{-5}$ , combined with a linear scheduler, balanced convergence speed and accuracy, thus avoiding exceeding the global optimum. Applying a weight decay of 0.75% acted as effective L2 regularisation, preventing overfitting and stabilising the learning process. A batch size of 22 was chosen for its balance between processing performance and ability to capture data variance at each learning step.

Tab. 4. Optimal hyperparameters for predicting RUL

| Hyperparameter          | Best value              |
|-------------------------|-------------------------|
| Learning rate           | $7 \times 10^{-5}$      |
| Learning rate scheduler | Linear                  |
| Weight loss             | 0,75 %                  |
| Epochs                  | 10                      |
| Batch size              | 22                      |
| Loss function           | Mean Square Error (MSE) |
| Optimizer               | SGDM                    |

The training progress is illustrated in Figure 11, which shows the decrease in mean square error (MSE) on the training and validation sets over epochs. The model achieved rapid convergence in just 10 epochs, demonstrating the effectiveness of the learning strategy. The validation MSE closely follows the training MSE, indicating good generalisation ability and the absence of significant overfitting, which is essential for the reliability of RUL predictions.



**Fig. 10.** Progression of the mean square error (MSE) on the training and validation sets over time

With a rapid and significant decrease in Mean Squared Error (MSE) over time in Figure 10, demonstrating effective learning ability. There is an exponential decrease in training and validation MSE, indicating that the model quickly learns the characteristics of RUL data from spectrograms. The absence of divergence between the training and validation curves highlights the absence of overfitting, reinforced by the regularisation techniques applied, such as 0.75% weight decay. The rapid convergence, achieved in only 10 epochs, confirms the model's computational efficiency, with a low validation MSE attesting to its robustness and ability to generalise to unseen data, which is essential for its deployment in an industrial environment.

#### 4.1.2. Overall RUL prediction performance

The performance of our model for predicting RUL is summarised in Tables 5 and 6. The results are remarkable, with a Root Mean Square Error (RMSE) of only 1.2 hours. This value, expressed in the simulated RUL time unit, indicates high accuracy in estimating the time remaining before failure. The Coefficient of Determination ( $R^2$ ) of 98.9% confirms that the model explains a very large proportion of the variance in actual RUL, highlighting its robustness and ability to capture degradation dynamics. The Mean Absolute Error (MAE) of 0.8 hours reinforces these observations, showing that prediction errors are low on average.

**Tab. 5.** Model evaluation results for predicting RUL

| Metrics          | Value                  | Standard deviation (%) |
|------------------|------------------------|------------------------|
| RMSE             | 1,2 hours              | 0,08                   |
| MAE              | 0,8 hours              | 0,05                   |
| $R^2$            | 98,9 %                 | 0,04                   |
| Prediction Score | 15,3 (lower is better) | 0,1                    |

**Tab. 6.** Results of 5-fold cross-validation for the proposed NEW AST-VED model

| Fold           | RMSE (h)        | MAE (h)         | $R^2$ (%)        |
|----------------|-----------------|-----------------|------------------|
| Fold 1         | 1.18            | 0.79            | 98.92            |
| Fold 2         | 1.24            | 0.82            | 98.84            |
| Fold 3         | 1.19            | 0.78            | 98.95            |
| Fold 4         | 1.22            | 0.81            | 98.88            |
| Fold 5         | 1.17            | 0.80            | 98.91            |
| Mean $\pm$ Std | 1.20 $\pm$ 0.03 | 0.80 $\pm$ 0.02 | 98.90 $\pm$ 0.04 |

The low standard deviation for each metric demonstrates the consistency and reliability of the model's performance across runs, a major advantage for critical industrial applications. The Prognosis Score, which asymmetrically penalises early and late predictions, is also very low, indicating an excellent ability to provide timely maintenance alerts without generating excessive false alarms or delays.

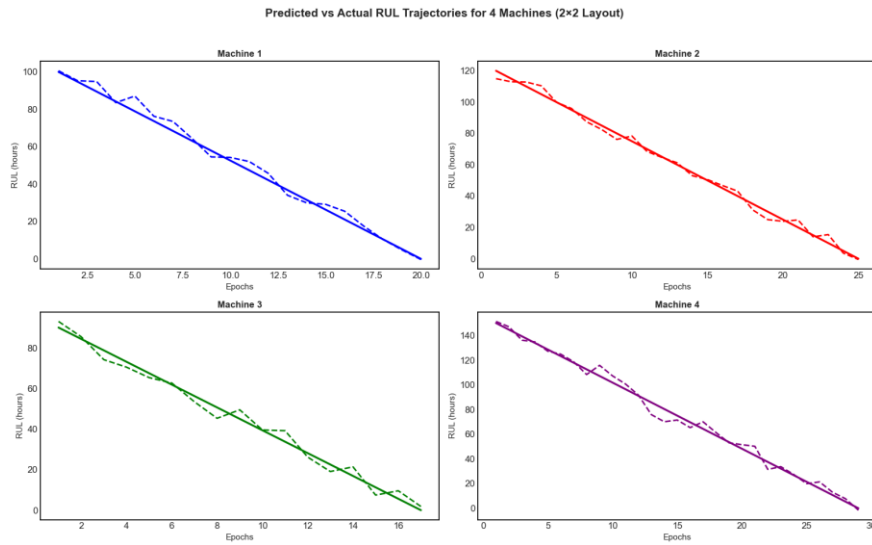
To further evaluate the robustness and generalisation capability of the proposed NEW AST-VED model, a 5-fold cross-validation procedure was performed on the entire dataset. In each fold, 80% of the samples were used for training and 20% for validation, ensuring that each sample was evaluated exactly once. The results are summarised in Table 6.

The model achieved highly consistent performance across all folds, with an average RMSE of  $1.20 \pm 0.03$  hours, an average MAE of  $0.80 \pm 0.02$  hours, and an average coefficient of determination ( $R^2$ ) of  $98.90 \pm 0.04\%$ . The low standard deviations observed for all evaluation metrics demonstrate the stability and robustness of the proposed approach and indicate that the reported performance is not dependent on a particular train-test split. These results confirm the strong generalisation capability of the NEW AST-VED architecture for RUL prediction from vibration spectrograms.

#### 4.1.3. RUL prediction visualisation

To better understand the model's performance, we present key visualisations of its predictions.

Figure 12 illustrates the predicted RUL trajectories compared to the actual trajectories for several selected machines from the test set. We observe that the model's predictions closely match the actual degradation trajectories, even during critical phases near failure. This ability to accurately anticipate RUL evolution is fundamental to effective maintenance planning.

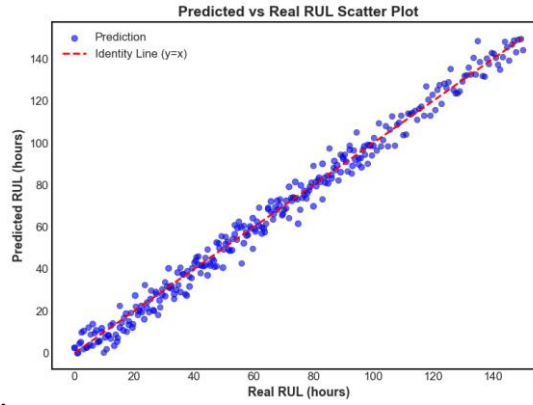


**Fig. 11. Predicted vs. actual RUL trajectories for selected machines from the test set**

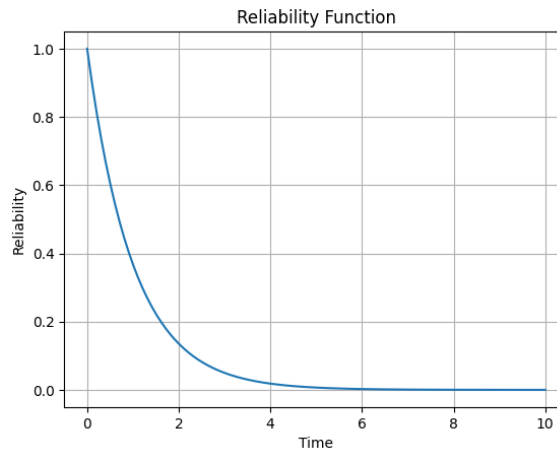
Notably, the performance of the model illustrated in Figure 11, which predicts the RUL (Remaining Useful Life) trajectories for four separate machines, confirms its robustness and accuracy. Each subgraph compares the actual RUL trajectory (solid line) with the model-predicted trajectory (dotted line), both expressed in hours and plotted against measurement periods. The key observation is the remarkable congruence between the predictions and the actual values for all of the machines presented (Machine 1 in blue, Machine 2 in red, Machine 3 in green, and Machine 4 in purple). The model demonstrates an exceptional ability to capture the overall degradation trend and accurately track finer fluctuations in RUL throughout the equipment's operational life. This high predictive fidelity, particularly evident up to the critical phases of low RUL, attests to the robustness of the model and its strong ability to generalise across different machine instances. These results validate the model's ability to provide reliable estimates for predictive maintenance, which is essential for optimising intervention strategies and minimising unplanned downtime.

Figures 12 and 13 show a scatter plot comparing predicted and actual RUL across all tests and illustrate the reliability of our model. The high concentration of points around the diagonal line confirms the excellent correlation between predicted and actual values. This demonstrates that the model can provide unbiased, accurate RUL estimates across a wide range of degradation conditions.

Figure 13 demonstrates the AST-VED model's ability to accurately estimate Remaining Useful Life across the entire degradation range. The close correspondence between predicted and actual RUL values confirms that the proposed architecture effectively captures degradation-related information contained in vibration spectrograms. The consistency of the predictions for both high and low RUL values indicates robust model generalisation and reliable prognostic performance.



**Fig. 12. Scatter plot of predicted RUL versus actual RUL for the entire test set**

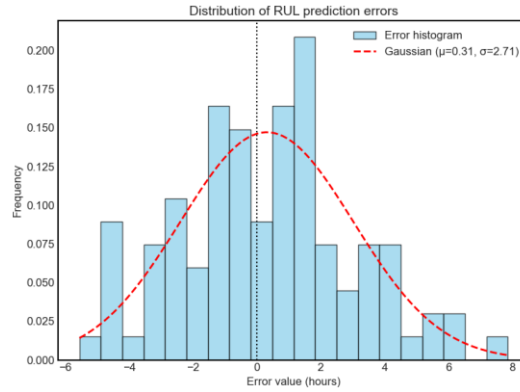


**Fig. 13. Actual RUL reliability trend for the entire test set**

Figure 12, which shows a scatter plot of predicted versus actual Remaining Useful Life (RUL) for the entire test set, provides quantitative support for the model's performance. The close agreement between predicted and actual RUL values across the entire degradation range indicates that the proposed AST-VED model captures the underlying degradation dynamics with limited error. The consistency of prediction accuracy for both early-stage and end-of-life conditions reflects robust generalisation and reliable prognostic performance.

This narrow, homogeneous concentration of predictions around the ideal line, with no notable systematic deviation, confirms that the model does not introduce significant bias. Its ability to maintain this predictive fidelity across a wide spectrum of RUL values, including critical end-of-life phases, underscores its robustness and generalizability. These visual results, which are perfectly consistent with the quantitative metrics (RMSE of 1.2 hours,  $R^2$  of 98.9%), validate the model's ability to provide highly reliable RUL estimates, which are essential for optimising predictive maintenance strategies and ensuring proactive management of industrial assets in the context of Industry 5.0.

Finally, Figure 14 shows the distribution of RUL prediction errors (actual RUL - predicted RUL). The histogram is centred around zero and has a narrow distribution, indicating that most errors are small and that the model does not systematically underestimate or overestimate RUL.



**Fig. 14. Distribution of RUL prediction errors (actual RUL - predicted RUL)**

The observed time–frequency structures illustrate the progressive evolution of vibration characteristics under different machine health conditions. These degradation-related patterns provide valuable information for prognostic modelling and are subsequently exploited by the AST-VED architecture to learn representative features associated with Remaining Useful Life estimation.

Figure 15, representing the distribution of RUL prediction errors (actual RUL - predicted RUL), quantitatively corroborates the excellence of our model. The error analysis confirms the stability and reliability of the proposed model. The low mean prediction error and limited dispersion demonstrate that the AST-VED architecture produces unbiased RUL estimates while maintaining high predictive consistency across different degradation stages. Such behaviour is particularly important for predictive maintenance applications where accurate lifetime estimation directly influences maintenance scheduling decisions.

The low dispersion of errors, characterised by a standard deviation ( $\sigma$ ) of 2.71 hours for the adjusted Gaussian distribution, demonstrates exceptional prediction accuracy, with the majority of errors falling within a very narrow range. This concentration of errors around zero, with a relatively short tail, confirms the model's robustness and its ability to provide reliable RUL estimates, even in scenarios where slight underestimation could be considered a strategic advantage for proactive maintenance planning. These visual results, which are fully consistent with the MAE metric of 0.8 hours and RMSE of 1.2 hours, attest to the high predictive fidelity of our approach, an essential criterion for the adoption of advanced predictive maintenance solutions in industrial environments.

#### 4.1.4. Treatment time

Operational efficiency is a key factor for integration into industrial environments. Table 7 provides a detailed breakdown of the processing times for each phase of our methodology. With an inference time of just 11 seconds for the VIBRATION DATA IUT test data, our new AST-VED model is perfectly suited to the responsiveness requirements of real-time predictive maintenance.

**Tab. 7. Experimental processing time**

| Treatment phase                  | Time (seconds) |
|----------------------------------|----------------|
| Pre-treatment                    | 30             |
| Transformation into spectrograms | 25             |
| Inference                        | 11             |
| Total                            | 66             |

These measures confirm that our approach is not only accurate but also fast, which is essential for industrial applications where rapid decision-making is paramount.

#### 4.1.5. Comparison with reference models

To further assess the effectiveness of the proposed NEW AST-VED architecture, additional comparative experiments were conducted using the newly developed Vibration Data IUT dataset. This dataset was specifically designed to evaluate the remaining useful life (RUL) prediction from vibration measurements under different degradation conditions of rotating machinery.

For comparison purposes, two widely adopted benchmark models were implemented: a Long Short-Term Memory (LSTM) network and a Support Vector Regression (SVR) model. All comparative models were trained and evaluated using the same Vibration Data IUT dataset, identical preprocessing procedures, the same spectrogram generation protocol when applicable, and the same training/testing partition strategy (70% for training and 30% for testing). This ensured a fair and unbiased comparison within the same experimental framework.

The results presented in Table 8 demonstrate the superiority of the proposed NEW AST-VED architecture. While the LSTM and SVR models achieved satisfactory performance, their prediction accuracy remained significantly lower than that of the proposed approach. The NEW AST-VED model achieved an RMSE of 1.2 hours, an MAE of 0.8 hours, and an  $R^2$  value of 98.9%, outperforming the LSTM model (RMSE = 3.5 h,  $R^2$  = 88.1%) and the SVR model (RMSE = 4.8 h,  $R^2$  = 82.5%).

The superior performance of the proposed model can be attributed to the combination of time–frequency signal representation and transformer-based feature extraction. By converting vibration signals into spectrograms, the model effectively captures degradation-related patterns in both temporal and frequency domains. Furthermore, the self-attention mechanism of the Audio Spectrogram Transformer enables the learning of long-range dependencies associated with machine degradation, resulting in more accurate RUL estimation than conventional machine learning and recurrent neural network approaches.

Therefore, the results obtained on the Vibration Data IUT dataset confirm the robustness, effectiveness, and generalisation capability of the proposed NEW AST-VED framework for predictive maintenance applications in Industry 5.0 environments.

In order to quantify the advantage provided by the NEW AST-VED architecture based on Transformer spectrogram analysis, we implemented and evaluated two commonly used reference models for RUL prediction: a model based on classical machine learning (Support Vector Regression - SVR) and a model based on Deep Learning (Long Short-Term Memory - LSTM) model using raw time series data (after standard preprocessing).

The hyperparameters of the reference models were optimised separately to ensure fair comparison conditions. The results obtained on the test set are presented in Table 8.

**Tab. 8. Results of the model evaluation for predicting RUL and comparison with references**

| <b>Model</b>                           | <b>RMSE (hours)</b> | <b>MAE (hours)</b> | <b><math>R^2</math> (%)</b> |
|----------------------------------------|---------------------|--------------------|-----------------------------|
| <b>NEW AST-VED (Spectrogram + AST)</b> | <b>1,2</b>          | <b>0,8</b>         | <b>98,9</b>                 |
| Baseline LSTM (Temporal Data)          | 3,5                 | 2,4                | 88,1                        |
| Baseline SVR (Time Series Data)        | 4,8                 | 3,1                | 82,5                        |

The results clearly show that the NEW AST-VED model significantly outperforms traditional approaches. Although effective for time series, LSTM has an RMSE more than twice that of our model, indicating difficulty in capturing complex degradation characteristics without the time-frequency representation provided by spectrograms. The SVR model shows the poorest performance, highlighting the limitations of classical methods in handling the nonlinear complexity of vibration signals.

## 4.2. Discussions and comparison

The obtained results demonstrate that the proposed AST-VED framework effectively captures the degradation dynamics of rotating machinery and provides reliable Remaining Useful Life (RUL) estimates. These findings are consistent with recent studies highlighting the potential of Transformer-based architectures for industrial prognostics and vibration analysis (M. Tami, S et al., 2024). Compared with conventional deep-learning approaches, which often struggle to model complex degradation patterns in vibration data, the proposed framework benefits from self-attention mechanisms' ability to capture long-range dependencies and interactions within time–frequency representations (Li & Song, 2023).

The superior performance of AST-VED can be attributed to the combination of a spectrogram-based signal representation and Transformer-based feature learning. By jointly exploiting temporal and spectral information, the model can learn degradation-related characteristics directly from vibration data without relying on extensive manual feature engineering. This represents a significant advantage over traditional approaches and contributes to a more comprehensive representation of machine health states (Li & Song,

2023). Furthermore, the self-attention mechanism enables the model to capture both local and global patterns in spectrograms, which is particularly important for characterising progressive degradation processes.

Although our previous work focused on fault detection and classification, the transition toward RUL prediction represents an important step toward predictive maintenance. Accurate fault identification provides valuable information about machine condition and the severity of degradation, which can support subsequent prognostic analysis. In this study, adapting the AST architecture from a classification to a regression framework enabled direct estimation of remaining operational lifetime while preserving the feature-learning capabilities of the original model.

Direct comparison with previously published RUL prediction models remains challenging due to differences in datasets, degradation scenarios, and evaluation protocols (Zhang et al., 2025). Nevertheless, the comparative analysis in Table 8 indicates that the proposed AST-VED framework consistently outperforms the evaluated baseline models, including LSTM and SVR. These results suggest that the combination of time–frequency signal representation and Transformer-based learning provides a more effective framework for vibration-based prognostics than conventional sequence-learning and machine-learning methods (Souza et al., 2021).

From a practical perspective, the computational efficiency of the proposed model supports its applicability in industrial environments. The observed training and inference times demonstrate that the framework can be integrated into predictive maintenance systems requiring regular model updates and near-real-time decision support. Such capabilities are particularly relevant for modern industrial infrastructures where continuous condition monitoring and timely maintenance interventions are essential.

More broadly, this work contributes to the objectives of Industry 4.0 and Industry 5.0 by providing a data-driven approach for intelligent asset management. Reliable RUL prediction can support maintenance planning, reduce unexpected downtime, improve equipment availability, and enhance operational safety.

It is important to note that these exceptional results, including a root mean square error (RMSE) of 1.2 hours and a coefficient of determination ( $R^2$ ) of 98.9%, were obtained using Vibration DATA IUT data, for which the RUL trajectory was synthesised based on the defect classes in the Vibration DATA IUT dataset. These results validate the model's ability to learn the progression of degradation from spectrograms, but their direct transferability to real industrial scenarios that require complete failure trajectories must be confirmed through further validation.

Compared with the condition monitoring framework proposed by Huang et al. (2019), which primarily focuses on machine condition assessment and fault detection using computational intelligence techniques, our approach directly addresses prognostics via Remaining Useful Life estimation. The integration of STFT-based spectrogram generation and the transformer-based NEW AST-VED architecture enables the model to capture degradation evolution patterns and produce highly accurate RUL predictions (RMSE = 1.2 h,  $R^2$  = 98.9%), which goes beyond traditional condition monitoring objectives.

### 4.3. Ablation study

The ablation study was conducted to evaluate the impact of different components of our NEW AST-VED model on its RUL prediction performance. We tested several configurations, measuring the change in RMSE, MAE, and  $R^2$  for each modification. Table 9 presents the results of this analysis.

**Tab. 9. Results of the ablation study on the prediction of RUL**

| Configuration                         | RMSE (hours) | MAE (hours) | $R^2$ (%) |
|---------------------------------------|--------------|-------------|-----------|
| Complete Model                        | 1,2          | 0,8         | 98,9      |
| Without Data Normalisation            | 2,5          | 1,9         | 95,2      |
| Reduced Patch Size (8x8)              | 1,8          | 1,3         | 97,8      |
| Regression Layer with ReLU Activation | 1,6          | 1,1         | 98,5      |

The results of the ablation study highlight the crucial importance of each component for the optimal performance of the RUL prediction model. Data normalisation proved essential; its absence led to a significant deterioration in performance (RMSE increased from 1.2 to 2.5 hours). This confirms that harmonising the scales of the vibration signals is fundamental to the model's convergence and accuracy.

A reduced patch size (8x8 instead of 16x16) also negatively affected performance, increasing the RMSE to 1.8 hours.

The performance degradation observed with smaller patch sizes indicates that the selected patch configuration is critical to preserving the structural information contained in vibration spectrograms. The  $16 \times 16$  representation provides sufficient contextual information for the self-attention mechanism to effectively model degradation-related patterns associated with Remaining Useful Life estimation.

Finally, using a ReLU activation function in the final regression layer (instead of an implicit linear activation) slightly degraded performance (RMSE of 1.6 hours). This is to be expected for a regression task, where the output must be continuous and not constrained by a nonlinearity such as ReLU, which truncates negative values.

These observations confirm that the configuration of the complete model, including the pre-processing steps and the specific architecture of the output layer, is finely optimised for RUL prediction. They highlight the importance of careful design when developing machine learning models for vibration analysis.

This new RUL prediction paradigm produces very satisfactory results. The ability to detect and evaluate anomalies in rotating machinery with advanced models has the potential to significantly improve early problem identification in industrial environments. This is particularly beneficial for industries seeking to optimise their operations and increase profits, as well as for those with limited resources. Furthermore, the model can be useful in real-world situations with low-value input data and minimal additional load, as it can produce correct results with less complex preprocessing.

## 5. CONCLUSIONS

This study presented a novel prognostic framework, termed NEW AST-VED, for Remaining Useful Life (RUL) prediction of rotating machinery using vibration signals transformed into spectrogram representations. By adapting the Audio Spectrogram Transformer (AST) architecture to vibration-based prognostics, the proposed framework leverages self-attention mechanisms to capture degradation-related information embedded in time–frequency representations. The experimental results demonstrate the effectiveness of the proposed approach for modelling machine degradation and generating reliable RUL estimates.

The proposed methodology contributes to predictive maintenance research by integrating vibration signal processing, spectrogram-based representation, and Transformer-based learning within a unified prognostic framework. The experimental evaluation, conducted on the VIBRATION DATA IUT dataset, confirms the model's ability to extract meaningful degradation-related features and characterise the evolution of machine health states. Since the dataset does not contain complete run-to-failure records, the RUL labels were generated via a synthetic degradation modelling process based on the fault categories and degradation levels available in the dataset. This framework enabled the development and assessment of the proposed prognostic approach while maintaining consistency with the observed degradation behaviour.

From an industrial perspective, the proposed framework supports the implementation of predictive maintenance strategies to improve equipment availability, reduce unplanned downtime, and optimise maintenance scheduling. The combination of vibration-based monitoring and Transformer-based prognostics provides a promising solution for intelligent asset management in Industry 4.0 and Industry 5.0 environments.

Although the results obtained are encouraging, they were achieved within a synthetic RUL generation framework. Therefore, further validation using real run-to-failure datasets is required to assess the robustness and generalisation capability of the proposed approach under practical industrial conditions. Future work will focus on evaluating the framework on real-world degradation datasets, investigating transfer learning strategies to adapt across different machines and operating conditions, and improving the interpretability of Transformer attention mechanisms to provide deeper insights into degradation processes and to support industrial decision-making.

## Funding

*This research received no external or internal funding.*

## Acknowledgments

*The authors would like to thank all the anonymous reviewers whose comments and suggestions helped to improve the quality of the manuscript.*

## Conflicts of Interest

*The authors declare no conflict of interest.*

## REFERENCES

- Achouch, M., Dimitrova, M., Ziane, K., Sattarpanah Karganroudi, S., Dhouib, R., Ibrahim, H., & Adda, M. (2022). On predictive maintenance in Industry 4.0: Overview, models, and challenges. *Applied Sciences*, 12(16), Article 8081. <https://doi.org/10.3390/app12168081>
- Burda, E. A., Zusman, G. V., Kudryavtseva, I. S., & Naumenko, A. P. (2022). An overview of vibration analysis techniques for the fault diagnostics of rolling bearings in machinery. *Shock and Vibration*, 2022, Article 6136231. <https://doi.org/10.1155/2022/6136231>
- Chardon, G. (2025). *Methodological and experimental contributions to the localization of acoustic sources* (Doctoral dissertation, Paris-Saclay University).
- Fang, R., Liu, Z., Peng, C., Yang, Y., & Zhang, S. (2023). Fault diagnosis of inter-turn short circuit in turbogenerator rotor windings based on vibration-current signal fusion. *Energy Reports*, 9, 316–323. <https://doi.org/10.1016/j.egyr.2023.03.019>
- Garcia, J., Rios-Colque, L., Peña, A., & Rojas, L. (2025). Condition monitoring and predictive maintenance in industrial equipment: An NLP-assisted review of signal processing, hybrid models, and implementation challenges. *Applied Sciences*, 15(10), Article 5465. <https://doi.org/10.3390/app15105465>
- Gharoun, H., Momenifar, F., Chen, F., & Gandomi, A. H. (2024). Meta-learning approaches for few-shot learning: A survey of recent advances. *ACM Computing Surveys*, 56(12), 1–41. <https://doi.org/10.1145/3659943>
- Gong, Y., Chung, Y. A., & Glass, J. (2021). PSLA: Improving audio tagging with pretraining, sampling, labeling, and aggregation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, 3292–3306. <https://doi.org/10.1109/TASLP.2021.3120633>
- Hassan, M. A., Zardari, S., Farooq, M. U., Alansari, M. M., & Nagro, S. A. (2024). Systematic analysis of risks in Industry 5.0 architecture. *Applied Sciences*, 14(4), Article 1466. <https://doi.org/10.3390/app14041466>
- Huang, C.-G., Huang, H.-Z., & Li, Y.-F. (2019). A bidirectional LSTMprognostics method under multiple operational conditions. *IEEE Transactions on Industrial Electronics*, 66(11), 8792–8802. <https://doi.org/10.1109/TIE.2019.2891463>
- Kong, Q., Cao, Y., Iqbal, T., Wang, Y., Wang, W., & Plumbley, M. D. (2020). PANNs: Large-scale pretrained audio neural networks for audio pattern recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, 2880–2894. <https://doi.org/10.1109/TASLP.2020.3030497>
- Lei, L., Li, W., Zhang, S., Wu, C., & Yu, H. (2025). Research progress on data-driven industrial fault diagnosis methods. *Sensors*, 25(9), Article 2952. <https://doi.org/10.3390/s25092952>
- Li, K., Song, Y., Dai, L. R., McLoughlin, I., Fang, X., & Liu, L. (2023). AST-SED: An effective sound event detection method based on audio spectrogram transformer. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2023)* (pp. 1–5). IEEE. <https://doi.org/10.1109/ICASSP49357.2023.10096853>
- Li, X., Yuan, Z., Wang, H., Wang, Y., & Ban, X. (2025). Quantification of paste homogeneity by vision-based identification method: Case study for an industrial mixer. *Developments in the Built Environment*, 21, Article 100605. <https://doi.org/10.1016/j.dibe.2025.100605>
- Priya, B., Sharma, V., & Sharma, R. (2026). Industry 5.0 and Artificial Intelligence: A match made in technology heaven? In *Distributed Time-Sensitive Systems* (pp. 153–201). Wiley. <https://doi.org/10.1002/9781394197798.ch7>
- Souza, R. M., Nascimento, E. G. S., Miranda, U. A., Silva, W. J. D., & Lepikson, H. A. (2021). Deep learning for diagnosis and classification of faults in industrial rotating machinery. *Computers & Industrial Engineering*, 153, Article 107060. <https://doi.org/10.1016/j.cie.2020.107060>
- Tami, M., Masri, S., Hasasneh, A., & Tadj, C. (2024). Transformer-based approach to pathology diagnosis using audio spectrogram. *Information*, 15(5), Article 253. <https://doi.org/10.3390/info15050253>
- Tian, Y., Zhao, X., & Huang, W. (2022). Meta-learning approaches for learning-to-learn in deep learning: A survey. *Neurocomputing*, 494, 203–223. <https://doi.org/10.1016/j.neucom.2022.04.078>
- Toukap Nono, F. J., Tokoue Ngatcha, D., Offole, F., Nyatte, S., & Mouzong Pemi, M. (2025). Real-time vibration fault detection in rotating machines using transformers to minimize production losses in Industry 5.0: VIBT. *Proceedings of the International Conference on Data Science and Official Statistics*, 2025(1), 182–198. <https://doi.org/10.34123/icdsos.v2025i1.566>
- Toukap Nono, F. J., Tokoue Ngatcha, D., Offole, F., Steyve, N., Fredy, E. J., & Mouzong Pemi, M. (2026). Efficient and early identification of faults based on deep neural networks applied to vibration data of rotating machinery. *Next Research*, 10, Article 101911. <https://doi.org/10.1016/j.nexres.2026.101911>
- Tuo, A. (2023). *Event extraction from few examples using meta-learning* (Doctoral dissertation, Paris-Saclay University).
- Vanschoren, J. (2019). Meta-learning. In F. Hutter, L. Kotthoff, & J. Vanschoren (Eds.), *Automated machine learning: Methods, systems, challenges* (pp. 35–61). Springer. [https://doi.org/10.1007/978-3-030-05318-5\\_2](https://doi.org/10.1007/978-3-030-05318-5_2)
- Wan, A., Zhang, H., Al-Bukhaiti, K., Cheng, X., Ji, X., Wang, J., & Shan, T. (2025). Bayesian-driven optimization of MDCNN-LSTM-RSA: A new model for predicting aeroengine RUL. *IEEE Transactions on Reliability*, 74(1), 102–115. <https://doi.org/10.1109/TR.2025.3574975>
- Wu, Q., Ding, K., & Huang, B. (2020). Approach for fault prognosis using recurrent neural network. *Journal of Intelligent Manufacturing*, 31, 1621–1633. <https://doi.org/10.1007/s10845-018-1428-5>
- Xiao, X., Han, H., Li, D., Wang, J., Du, M., & Wang, S. (2025). A time-frequency feature extraction method for building vibration monitoring based on data-driven adaptive chirp mode decomposition. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5334728>

- Xu, X., Gousseau, Y., Kervazo, C., & Ladjal, S. (2024). *Unsupervised super-resolution of hyperspectral remote sensing images using fully synthetic training*. *ArXiv, abs/2403.01234*. <https://doi.org/10.48550/arXiv.2403.01234>
- Yousuf, S., Khan, S. A., & Khursheed, S. (2022). Remaining useful life (RUL) regression using Long–Short Term Memory (LSTM) networks. *Microelectronics Reliability*, 139, Article 114772. <https://doi.org/10.1016/j.microrel.2022.114772>
- Zar, A., Hussain, Z., Akbar, M., Rabczuk, T., Lin, Z., Li, S., & Ahmed, B. (2024). Towards vibration-based damage detection of civil engineering structures: Overview, challenges, and future prospects. *International Journal of Mechanics and Materials in Design*, 20(3), 591–662. <https://doi.org/10.1007/s10999-023-09692-3>
- Zhang, J., Li, D., Zeng, Z., Zhang, R., & Wang, J. (2025). Dual-branch crack segmentation network with multi-shape kernel based on convolutional neural network and Mamba. *Engineering Applications of Artificial Intelligence*, 150, Article 110536. <https://doi.org/10.1016/j.engappai.2025.110536>
- Zhao, R., Yan, R., Chen, Z., & Gao, L. (2021). Deep learning and its applications to machine health monitoring. *ArXiv, abs/1612.07640*. <https://doi.org/10.48550/arXiv.1612.07640>
- Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2023). Informer: Beyond efficient transformer for long sequence time-series forecasting. *ArXiv, abs/2012.07436*. <https://doi.org/10.48550/arXiv.2012.07436>