Estimating Spall Severity in Rolling Element Bearings: A Supervised Learning Approach With Naturally Progressing Spalls

Stephan Baggeröhr[1,2], Cees Taal[1], and Konstantinos Gryllias[2,3]

[1] *SKF, Houten, Utrecht, The Netherlands*
stephan.baggerohr@skf.com
cees.taal@skf.com

[2] *KU Leuven, Department of Mechanical Engineering, Celestijnenlaan 300, 3001, Leuven, Belgium*
stephan.baggeroehr@kuleuven.be
konstantinos.gryllias@kuleuven.be

[3] *Flanders Make@KU Leuven, Belgium*

Abstract

Estimating the severity of localized defects in rolling element bearings is critical for accurate Remaining Useful Life (RUL) estimation, yet it remains challenging under non-stationary operating conditions with fluctuating speeds. Existing data-driven methods struggle to generalise due to the lack of high-fidelity, damage-progression data, and susceptibility to machine-specific structural transfer functions. A Siamese Transformer based neural network is utilized to predict continuous spall size directly from concurrent vibration measurements across three selected operating speeds, reducing the need for long-term trending. Using the amplitudes at the ball-pass frequency and its harmonics from spectra, and a data augmentation strategy, the proposed approach aims to decouple the fault signature from the system transfer function. Trained on a single run-to-failure dataset of one N209 ECP bearing with automated ground truth sizing for labels, the network acts on a regression target to learn the mapping between spectral features and the defect size. Preliminary results suggest that this proof-of-concept framework shows promise for generalization to unseen speed combinations and synthetic transfer function profiles within the scope of the studied experiment.

1. Introduction

Vibration-based monitoring remains essential for fault detection and Remaining Useful Life (RUL) prediction in rotating machinery. Bearings often continue to operate for extended periods after spall initiation while exhibiting a largely deterministic fault progression (Kotzalas & Harris, 2001; Rosado, Forster, Thompson, & Cooke, 2009; Mason, Trivedi, & Rosado, 2017). The ultimate objective is the precise estimation of this RUL. However, the realization of a generalizable RUL models is limited by the gap between laboratory training data and the complex, non-stationary realities of industrial deployments. While traditional data-driven deep learning models achieve high classification accuracies on discrete, artificially induced defects (such as electrical discharge machined notches), they frequently fail to capture the complex mechanical realities of naturally progressing fatigue spalls (Zhang et al., 2021). Furthermore, these models fundamentally struggle to generalize across different machines or test-rigs, exposing a bottleneck at the intersection of vibration-based condition monitoring and applied deep learning (Liefstingh, Taal, Restrepo, & Azarfar, 2021).

Towards domain generalization, a diagnostic model must isolate the physical signature of the defect from the signature of the machine it runs on as well as the stochastic environmental noise. The foundational premise of this tracking relies on the kinematic interaction between a rolling element and a spalled raceway. When a rolling element traverses a spall, the rapid de-stressing and subsequent re-stressing impact generates a continuous, periodic pulse train in the time domain (Bublil et al., 2025). Ap-

plying the Fourier transform to this pulse train yields a spectrum bounded by a characteristic sinc function envelope (Bublil et al., 2025; Bublil, Taal, Maljaars, Klein, & Bortman, 2024; Ngo-The, Taal, & Antoni, 2025). As the physical spall elongates due to rolling contact fatigue, the time-domain impact duration increases, systematically compressing the main lobe of the bounding sinc function. Consequently, while the absolute amplitude of the vibration signal is heavily influenced by external variables such as radial load and sensor damping, the relative ratios of the harmonic magnitudes (e.g., the first four harmonics of a specific fault frequency) are theoretically dictated strictly by the physical width of the spall (Ngo-The et al., 2025). To exploit this relationship, Ngo The et al. (Ngo-The et al., 2025) introduced a speed-normalized condition indicator designed to systematically eliminate the influence of the structural transfer function. Their analytical method achieves this cancellation by evaluating the amplitudes of corresponding fault harmonics across operating speeds that form strict integer speed-harmonic ratio pairs.

The measured acceleration from a spalled bearing operating in a machine, is the cross-spectral product of the underlying fault force and the structural transmission path. This path acts as a frequency-dependent filter, selectively amplifying fault frequencies that align with structural resonances and attenuating those at anti-resonances (Shen et al., 2021). The diagnostic complexity escalates under variable or non-stationary operating speeds. As the rotational frequency fluctuates, the fundamental fault harmonics dynamically sweep across the stationary resonance peaks of the transfer function, causing substantial amplitude modulation (Chen et al., 2025).

High-capacity deep learning algorithms, such as standard Deep Convolutional Neural Networks (CNNs), overfit to this amplitude modulation. Rather than learning the universal physical signatures of bearing degradation, these networks essentially memorize a lookup table of the training rig's specific resonance profile (Ni, Ji, Halkon, Feng, & Nandi, 2023). Consequently, when the model is deployed on a different physical machine, the resonance peaks shift, and the diagnostic accuracy of the purely data-driven model degrades markedly (Yang, Yang, Jin, & Liu, 2024).

Recent literature explicitly acknowledges this bottleneck. While some approaches attempt to dynamically fuse explicit speed information into non-stationary feature extraction architectures (Chen et al., 2025) , or embed physics-motivated loss functions into residual networks to force physical consistency (Shen et al., 2021; Ni et al., 2023), the mathematical disentanglement of the forcing function from the structural transmission path remains an ongoing challenge. Furthermore, the application of arbitrary mathematical transformations (e.g., standard data scaling) for deep learning data augmentation destroys the underlying physical meaning of vibration time-series data (Dong, Li, Zheng, Wang, & Xu, 2022). To achieve cross-machine generalization, augmentation strategies must be intrinsically grounded in analytical mechanical realities (Fan et al., 2025; Xu et al., 2023).

As a step toward addressing these limitations, a supervised learning framework is proposed that estimates continuous spall severity by synthetically augmenting training data with diverse, randomly generated structural transfer functions. This data is processed by a Siamese Transformer-based neural network designed to learn invariant mappings from harmonic amplitudes. By operating on four selected harmonics sampled concurrently from three arbitrary operating speeds, the framework reduces the need for explicit operational speed inputs or the characterization of static spectral patterns.

The Siamese architecture was a deliberate design choice driven by the physical constraints of the problem. Standard feed-forward networks, when presented with bivariate features (such as log-amplitude and frequency pairs), are highly susceptible to overfitting the training distribution, effectively resulting in a rigid lookup table. Conversely, Siamese networks are characterized by parallel branches utilizing shared weights. This makes Siamese networks intrinsically designed for metric learning and extracting relational invariants rather than absolute feature mappings (Bromley, Guyon, LeCun, Säckinger, & Shah, 1993; Chopra, Hadsell, & LeCun, 2005; Koch, Zemel, & Salakhutdinov, 2015). By independently processing the harmonic data from different operational speeds through identical, shared transformations, the network is forced to map the inputs into a common latent space. This architectural constraint prevents the model from blindly memorizing raw frequency-to-spall-size pairs; instead, it enforces the learning of the underlying physical relationship: the invariant amplitude ratios that define the progressing spall size independently of absolute speed.

The core novelty of this work lies in forcing the network to abandon absolute amplitude memorization in favor of contrastive metric learning (Vu, Nguyen, Tran, Pham, & Lo, 2024; Yang et al., 2024). By mathematically normalizing the data to cancel static transmission effects, and subsequently augmenting the training dataset with thousands of synthetic, second-order resonance responses, the network is exposed to a diverse but bounded array of synthetic transmission paths, parameterized by specific uniform distributions. Processed through the Siamese Transformer architecture, the model is encouraged to reduce sensitivity to transient, high-energy resonant ar-

tifacts, with the aim of computing metrics based on domain-invariant amplitude ratios across the multiple operating speeds (Song & Liu, 2024), capturing the underlying sinc function and mapping it to the geometric kinematics of the spall.

The main contributions of this work are summarized as follows:

1. A transfer function augmentation strategy that provides preliminary evidence that deep learning models may generalize to unseen synthetic transfer function profiles, as demonstrated in a proof-of-concept experiment on a single bearing type (N209 ECP).
2. The relaxation of strict integer speed ratio requirements, necessitating only three concurrent operating speeds for inference within the tested experimental setup, without requiring explicit speed input to the model.
3. Instantaneous regression directly to the physical size of the defect as demonstrated in the studied experimental setting.

The remainder of this paper is organized as follows. Section 2 summarizes the underlying bearing fault model and the theoretical foundation of harmonic amplitude analysis. Section 3 presents the proposed methodology, including the dataset, data normalization, symmetry-breaking target transformation, transfer function augmentation, and the Siamese Transformer neural network architecture. Section 4 reports the experimental results, demonstrating the generalization capability across unseen speeds and transfer functions. Section 5 discusses the influence of speed selection on prediction accuracy and the role of the Speed Ratio Span metric. Finally, Section 6 concludes the paper and outlines directions for future research.

2. Summary of the Bearing Fault Model

The baseline bearing fault model (Bublil et al., 2025) assumes that the underlying force signal $f(t)$ can be approximated as a periodic pulse wave. Let L denote the distance between two rolling elements (REs) and d the spall size. Expanding this force pulse into a Fourier series with a harmonic index $k \in \{1, 2, 3, \ldots\}$ yields the force coefficients f_k, which are given by a sinc function (see also (Ngo-The et al., 2025)):

$$f_k = \sin\left(\frac{k\pi d}{L}\right)\frac{1}{\pi k} = \mathrm{sinc}\left(\frac{kd}{L}\right)\frac{d}{L}. \tag{1}$$

As shown, f_k is independent of the shaft rotational speed Ω (expressed in RPM throughout this paper) and is based solely on the relative geometry of the fault d/L. Examples of the first four harmonics as a function of the normalized spall size are illustrated in Figure 1. As visible in the figure, each harmonic f_k traces a characteristic sinc-modulated amplitude pattern: the higher harmonics complete their oscillation cycles at progressively smaller values of d/L, and successive zero-crossings shift inward with increasing harmonic order. This produces a rich set of complementary, size-sensitive features that together uniquely characterize the spall geometry across the normalized size range. To estimate fault severity, Bublil et al. (Bublil et al., 2025) propose tracking these harmonic maxima and minima and mapping them to spall sizes. However, in practice, f_k cannot be measured directly. Instead, the spectrum of the acceleration response of the bearing to the fault frequency amplitude $c_k(\Omega)$ is analyzed. This measured amplitude is heavily modulated by the structural transfer function of the machine H:

$$c_k(\Omega) = H(k\Omega) f_k. \tag{2}$$

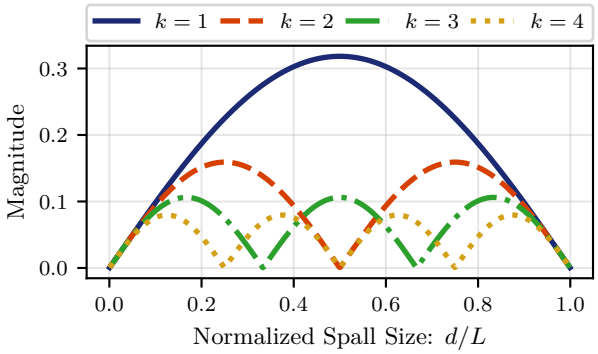

Figure 1. Bearing fault model Fourier magnitudes at the fault frequencies of a progressing outer ring spall.

Consequently, under fluctuating operating speeds, the signal extrema are distorted by the transfer function, preventing a direct mapping of amplitude trends to spall sizes. To get around this, Ngo The et al. (Ngo-The et al., 2025) proposed analyzing specific integer speed ratios to cancel out the transfer function entirely:

$$\frac{c_2\left(\frac{\Omega}{2}\right)}{c_1(\Omega)} = \frac{f_2}{f_1}. \tag{3}$$

However, this method is restricted to integer speed ratios and still relies on an undefined mapping from the coefficient ratio f_2/f_1 to the physical spall size.

3. Methodology

To accurately estimate spall severity without relying on strict integer speed ratios or historical trend tracking, a Siamese neural network framework trained on augmented

harmonic features is proposed. The following sections detail a data augmentation strategy designed to simulate a diverse array of structural transfer functions, H, and the architecture of the Siamese network. An overview of the complete five-stage methodological workflow is illustrated in Figure 2.

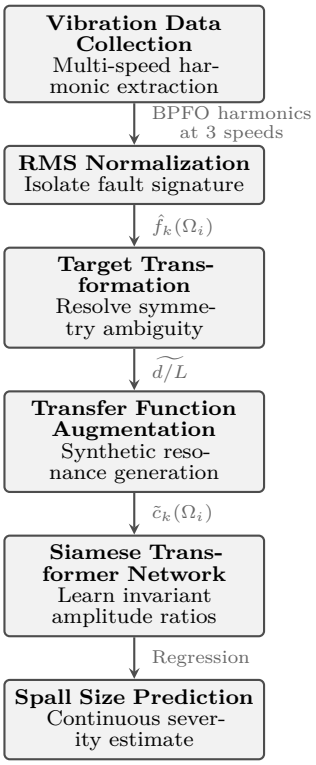

Figure 2. Overview of the proposed five-stage methodology.

The workflow begins with multi-speed vibration data collection and RMS normalization to isolate the fault signature from machine-specific transfer functions. A symmetry-breaking target transformation is then applied to resolve the inherent model ambiguity, after which the data are augmented with synthetic transfer functions to expose the network to diverse structural resonances. Finally, the Siamese Transformer network performs direct regression to predict a continuous spall size estimate. Each of these stages is described in detail in the following subsections.

3.1. Dataset

The dataset encompasses acceleration recordings captured from a naturally propagating bearing fault experiment, which was previously utilized by Bubil et al. (Bublil et al., 2024) and Ngo-The et al. (Ngo-The et al., 2025). An initially seeded small outer ring defect on an N209 ECP bearing was run-to-failure under high load. During intermediate progression intervals, snapshots were recorded across 10 different operating speeds ranging from 300 RPM to 3000 RPM. Ground truth spall tracking was governed by dynamic load-cell measurements validated against visual inspections (Bublil et al., 2024). The raw time-domain acceleration signals are acquired alongside concurrent rotational speed data. Minor speed fluctuations are mitigated with angular resampling, effectively mapping the signal from the time domain into the angular domain. Subsequently, the Power Spectral Density (PSD) of the resampled signal are calculated to represent the vibration energy across the frequency spectrum. Finally, peak detection is performed at the theoretical Ball Pass Frequency Outer race (BPFO) and its subsequent harmonics, up to and including the 4th harmonic.

3.2. Data normalization

A challenge in deep learning for bearing fault estimation is generalizing to machines with varying structural transfer functions, H. While the available training data is tied to the specific transfer function of the experimental setup, a robust model must operate seamlessly across new machines with distinct resonance characteristics.

The training dataset contains numerous observations of $c_k(\Omega_i)$ for normalized spall sizes ranging from 0 to 1, measured across multiple operating speeds Ω_i, where $i \in \{1, 2, \ldots, M\}$ and M is the total number of distinct operating speeds in the training dataset (here $M = 10$, corresponding to the ten speed setpoints from 300 to 3000 RPM described in Section 3). To approximate the underlying fault force f_k and reduce the influence of the transfer function of the experimental setup, the observed fault magnitudes are normalized by their Root Mean Square (RMS) value, calculated per speed and per harmonic:

$$\hat{f}_k(\Omega_i) = \frac{c_k(\Omega_i)}{\sqrt{\frac{1}{N}\sum_{j=1}^{N} c_k^2(\Omega_i, s_j)}}. \tag{4}$$

Here, N is the total number of spall size observations s_j in the training dataset, $c_k(\Omega_i, s_j)$ is the measured amplitude of the k-th harmonic at speed Ω_i for spall size s_j, and $\hat{f}_k(\Omega_i)$ is the normalized magnitude approximating a scaled version of f_k. As illustrated in the top panel of Figure 3, this normalization procedure approximates the theoretical model presented earlier in Figure 1. This step isolates the generalized fault signature, enabling the dataset to be augmented with diverse, synthetic transfer functions $\tilde{H}$ (defined below) during training.

3.3. Symmetry Breaking Target

To train the neural network effectively, the inherent mathematical symmetry in the underlying bearing fault model must be addressed. Because the harmonic magnitude responses are mirrored around the 50% mark of the rolling element distance, a model cannot intrinsically distinguish between spall sizes where $d/L < 0.5$ and those where $d/L > 0.5$. To resolve this ambiguity, the regression target is constrained:

$$\widetilde{d/L} = \min\left(\frac{d}{L}, 1 - \frac{d}{L}\right) \tag{5}$$

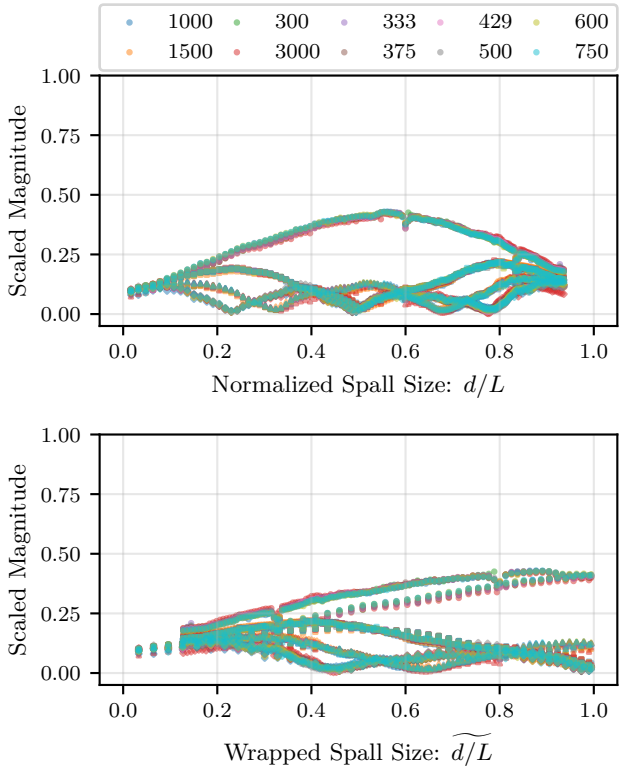

Figure 3. Normalized BPFO spectral amplitudes for the first four harmonics across all speeds (top), and symmetry-breaking regression target mapping (bottom).

The top plot of Figure 3 shows the normalized BPFO spectral amplitudes for the first four harmonics across all speeds, demonstrating the elimination of the effect of speed. The bottom plot shows the spectral amplitudes wrapped symmetrically around the 50% threshold ($d/L = 0.5$) to eliminate the inherent ambiguity of the sinusoidal patterns.

3.4. Data Augmentation via Transfer Functions

To improve the robustness of the model to unseen machine structures, synthetic transfer functions, denoted as $\tilde{H}$, are generated. These are formulated as standard second-order systems featuring randomly localized resonance peaks. Specifically, each synthetic transfer function is constructed by superimposing either one or two resonance peaks onto a baseline response. For each resonance r, three parameters are sampled from uniform distributions (where $\mathcal{U}(a, b)$ denotes the continuous uniform distribution over $[a, b]$): the resonance frequency $\nu_0^{(r)} \sim \mathcal{U}(50, 1000)$ Hz, the peak gain $G_{\text{peak}}^{(r)} \sim \mathcal{U}(0, 5)$ dB, and the quality factor $Q^{(r)} \sim \mathcal{U}(5, 20)$.

The frequency response magnitude of each individual resonance is modeled as:

$$\tilde{H}_r(\nu) = \frac{G_{\text{peak}}^{(r)}}{\sqrt{\left(1 - \left(\frac{\nu}{\nu_0^{(r)}}\right)^2\right)^2 + \left(\frac{\nu}{Q^{(r)}\nu_0^{(r)}}\right)^2}} \tag{6}$$

where $\nu = k \cdot \text{BPFO} \cdot \frac{\Omega}{60}$ represents the frequency (in Hz) of the k-th harmonic at a rotational speed of Ω (RPM). The total transfer function amplitude, expressed in the logarithmic domain (dB), is then computed by aggregating the baseline and resonance components:

$$\tilde{H}_{\text{total,dB}}(\nu) = H_{\text{baseline}} + \alpha \cdot \nu + \sum_{r=1}^{N_{\text{res}}} 20 \log_{10}(\tilde{H}_r(\nu)) \tag{7}$$

where $H_{\text{baseline}} \sim \mathcal{U}(-10, 0)$ dB represents a constant baseline attenuation, $\alpha \sim \mathcal{U}(-0.001, 0.001)$ dB/Hz introduces a slight frequency-dependent slope, and $N_{\text{res}} \in \{1, 2\}$ is the total number of simulated resonance peaks. By combining the normalized force magnitude from Equation 4 with the synthetic transfer function from Equation 7, the augmented observations are generated in logarithmic form:

$$20 \log_{10}(\tilde{c}_k(\Omega_i)) = \tilde{H}_{\text{total,dB}}\left(k \cdot \text{BPFO} \cdot \frac{\Omega_i}{60}\right) + 20 \log_{10}(\hat{f}_k(\Omega_i)). \tag{8}$$

Here, $\tilde{c}_k(\Omega_i)$ denotes the synthetic augmented harmonic amplitude. That is, an estimated acceleration response at the k-th BPFO harmonic and speed Ω_i after applying the randomly drawn transfer function $\tilde{H}$. $\tilde{c}_k(\Omega_i)$ is expressed in linear units before the logarithmic scaling.

As conceptualized in Figure 4, multiplying the normalized force estimates by a diverse array of these synthetic transfer functions exposes the neural network to heavily skewed resonance environments. During training, this data augmentation forces the model to learn the invariant ratio structures connecting the harmonics, rather than overfitting to the specific dynamics of the experimental test rig.

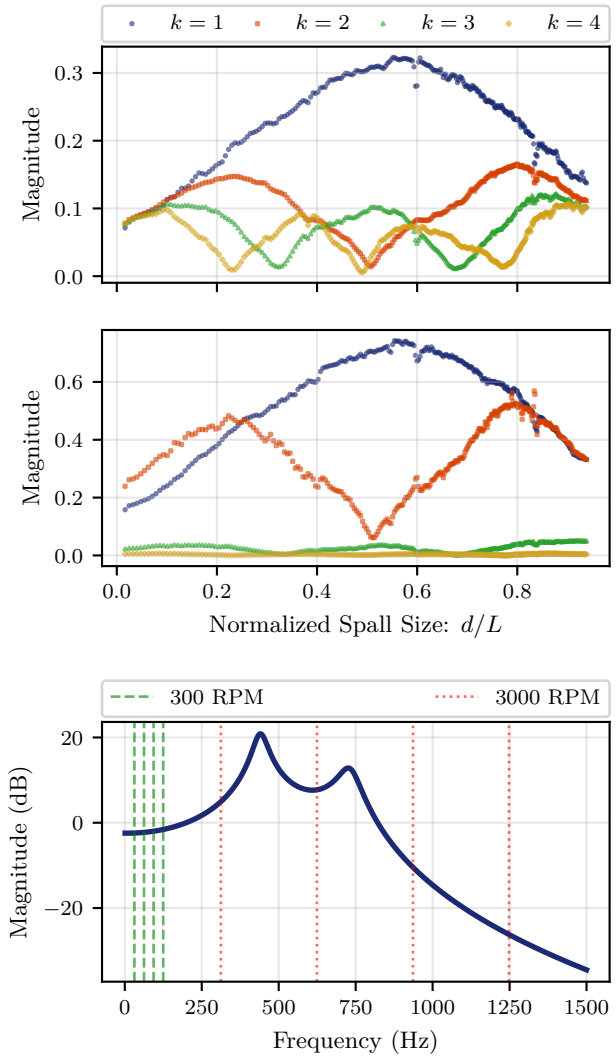

Figure 4. Synthetic transfer function ($\tilde{H}$) (bottom) augmentation applied to vibration data at 300 RPM (top) and 3000 RPM (middle).

In Figure 4, the generated log-domain resonance (bottom) is applied to distort the observed harmonic amplitudes at both 300 RPM (top) and 3000 RPM (middle), training the network to generalize across unseen host structures. In the 3000 RPM case, the first two harmonics lie within the system's resonance frequencies, causing their amplitude to be much higher than the last two.

3.5. Neural Network Architecture

As illustrated in Figure 5, the proposed model is a Siamese Transformer-based framework designed to process vibration harmonic data from three arbitrary machine speeds simultaneously.

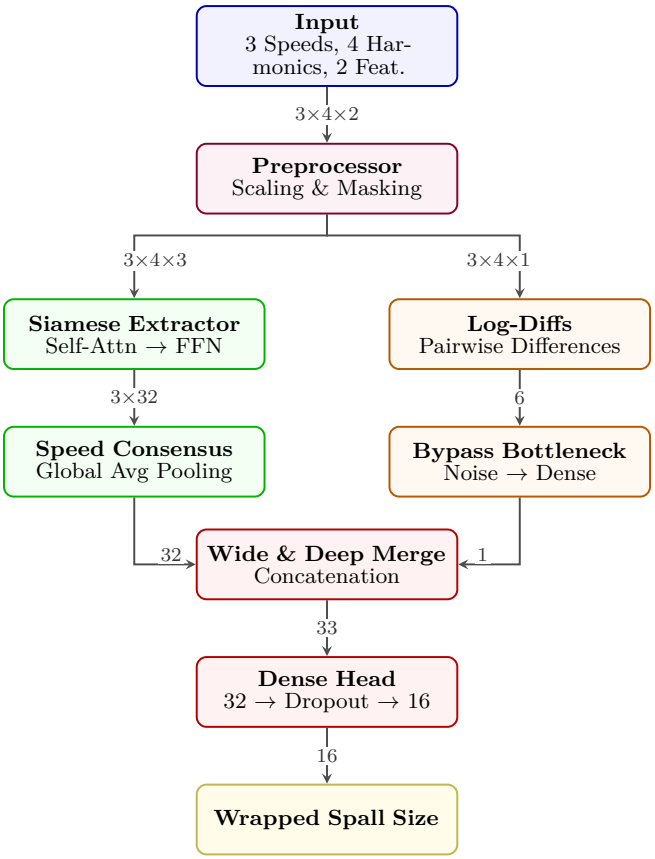

Figure 5. Architecture of the proposed Siamese Transformer-based neural network.

The model ingests a $3 \times 4 \times 2$ input matrix containing log-amplitudes and frequencies for four harmonics across three concurrent operating speeds. The data is processed through two parallel branches: a Shared Siamese Extractor that builds a consensus embedding from speed-specific interactions, and a Speed-Agnostic Log-Diff Bypass that explicitly extracts speed-invariant amplitude ratios. The extracted features are concatenated in a Wide & Deep Merge to predict the continuous, symmetry-breaking spall severity target.

The data flow begins with the input features, $\mathbf{C}$, which form a $3 \times 4 \times 2$ matrix representing:

$3\times$ operating speeds, $\Omega_i \in \{\Omega_1, \Omega_2, \Omega_3\}$,
$4\times$ BPFO harmonics, $k \in \{1, 2, 3, 4\}$,
$2\times$ features: the log-amplitude $(20 \log_{10}(c_k(\Omega_i)))$ and

the frequency $(k \cdot \text{BPFO} \cdot \frac{\Omega_i}{60})$ in Hz.

Before passing these input features into the parallel network branches, the log-amplitudes and frequencies are conditioned using a custom preprocessing layer. This layer performs three critical functions to ensure the network learns robust and physically meaningful representations:

- **Feature Scaling:** To prevent features with larger numerical magnitudes from dominating the learning process, the layer scales frequencies and amplitudes to specific physical bounds. Rather than absolute extremes, the logarithmic amplitudes undergo a robust min-max scaling bounded by the 1st and 99th percentiles of the training data (predefined as -5.0 to 8.0). This explicitly prevents extreme outliers from skewing the normalized distribution. The raw frequency values are normalized by dividing them by the maximum expected physical frequency (1250 Hz) of the setup.
- **Harmonic Positional Encoding:** Standard self-attention mechanisms are permutation-invariant and lack an inherent understanding of sequence order. To resolve this, a continuous positional encoding layer, comprising the values $[-1.0, -0.333, 0.333, 1.0]$, is concatenated to the scaled amplitude and frequency features. This ensures that the downstream Transformer blocks retain the structural identity of which of the four harmonics they are currently processing.
- **Harmonic Masking Strategy:** To improve the generalization of the model and act as a regularizer, a harmonic dropout mask is applied exclusively during the training phase. The layer utilizes a 15% masking probability to completely zero out the scaled amplitude and frequency values for randomly selected harmonics within a given batch. This forces the network to learn the invariant ratio structures connecting the surviving harmonics, rather than overfitting to a single dominant frequency peak.

Following preprocessing, the data is routed through two parallel pathways. The first is a *Shared Siamese Extractor*, where each branch processes one of the three speeds independently. This is achieved using self-attention mechanisms, Feedforward Neural Network (FFN) blocks, and Global Average Pooling (GAP). The network utilizes GAP to compress the four harmonics into a 32-dimensional Speed Fingerprint. A consensus layer then merges these three individual fingerprints into a single robust embedding.

In parallel with the Transformer branch, a *Speed-Agnostic Log-Diff Bypass* explicitly calculates pairwise log-amplitude differences among the four harmonics for each speed group. As demonstrated by Dang et al. (Ngo-The et al., 2025), diagnostic information is embedded in the amplitude ratios of bearing fault harmonics. Because the amplitudes are in the logarithmic domain, taking their difference mathematically isolates these invariant ratios, independent of the operating speed Ω. Finally, this bypass output is concatenated with the Transformer embedding and passed through a dense prediction head to output a continuous scalar value representing the target spall size.

4. Results

To evaluate the Siamese network's ability to generalize to unseen speed combinations and synthetic structural transfer functions, the model was trained on a fixed triplet of operating speeds and tested against the remaining speeds. The training and testing datasets featured entirely distinct sets of synthetic transfer functions. This process was repeated, retraining and evaluating the model from scratch, for all of the 120 unique 3-speed combinations available in the dataset.

The results provide evidence of the ability of the model to estimate spall dimensions within the tested experimental conditions. As shown in Figure 6, the predicted spall sizes, aggregated across all 120 speed triplets, correlate with the physical ground truth for the single N209 ECP bearing used in this study, throughout the increasing size of the fault.

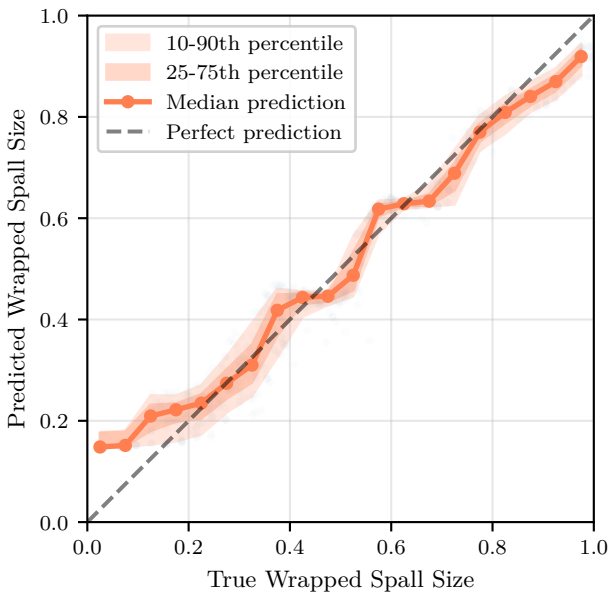

Figure 6. Predicted spall size evaluations against the load-cell validated ground truth over the naturally progressing fault timeline.

The envelopes in Figure 6 show the mean and percentile spread of predictions across all 120 speed combination models, tracking the fault progression over the sequentially recorded machine run.

While the overall aggregated performance of proposed method shows strong predictive capability with a low average error, a deeper analysis of the individual 3-speed subsets reveals significant variability in cross-speed transferability. This variance is underscored by the ranked distribution of the Mean Absolute Error (MAE) across the 120 unique combinations, shown in Figure 7. Certain speed triplets exhibit greater predictive stability than others when deployed against untested speeds and structural transfer functions.

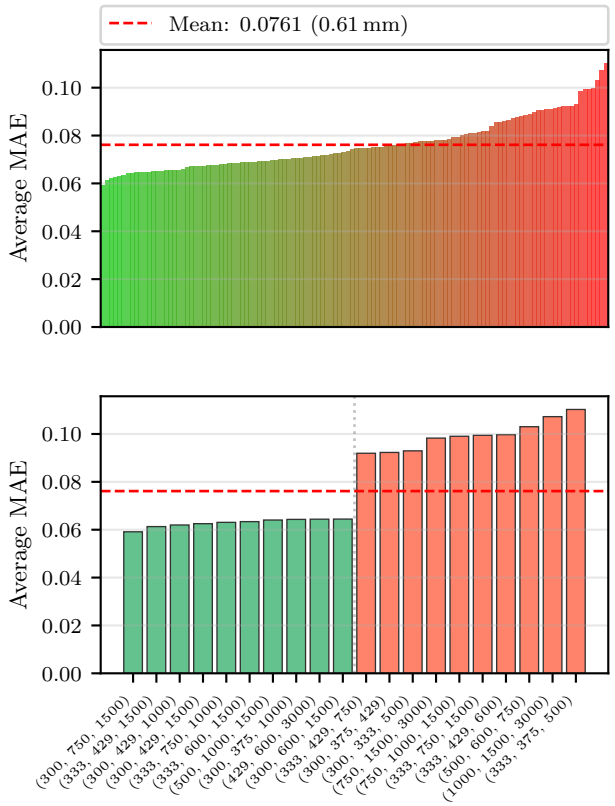

Figure 7. Distribution of Mean Absolute Error (MAE) across 120 unique 3-speed inference subsets (top), including a detailed view of the best and worst-performing speed triplets (bottom).

The ranked distribution in Figure 7 reveals that while the mean MAE across all 120 speed triplets is 0.076 in normalized units, the spread of the distribution is substantial: the best-performing triplets achieve markedly lower errors while the worst-performing subsets exhibit significantly higher MAE, as visible from both panels of Figure 7. Nevertheless, the mean MAE of 0.076 corresponds to an absolute spall sizing uncertainty of approximately 0.61 mm for the N209 ECP bearing geometry studied here, given an estimated rolling-element spacing L of approximately 16 mm. Given that practically relevant spall dimensions are expected to span several millimetres before reaching a critical contact geometry, an average absolute error on the order of 0.61 mm is considered small relative to the severity range of interest for early-to-mid-stage progression monitoring. At later stages, where precise sizing is most critical for RUL estimation, the error budget becomes more consequential and motivates further investigation.

The variability seen in Figure 7 is not solely attributable to random noise; it appears to be systematically driven by the selected speed combination, as subsequently formalized through the S_{span} metric, defined as:

$$S_{span} = \frac{\Omega_{max}}{\Omega_{min}} \tag{9}$$

where Ω_{max} and Ω_{min} represent the highest and lowest operating speeds within a given inference triplet, respectively. This metric quantitatively describes the "width" of the frequency coverage captured by the concurrent inputs.

As illustrated in Figure 8, evaluating the Mean Absolute Error (MAE) against S_{span} for all 3-speed combinations reveals a trend: expanding the inference speed span corresponds to a decrease in prediction error, yielding lower MAE and improved consistency.

A wider, more exhaustive speed ratio span correlates with lower prediction error and more stable generalization when the model is deployed alongside unknown machine transfer functions.

5. Discussion

The recorded outcomes suggest that tracking spall progression using a Siamese network, when mapped directly to normalized harmonic features, may reduce the scaling vulnerabilities inherent in raw vibration spectra, within the scope of the single-bearing experiment conducted here.

As established, the underlying fault force f_k responsible for bearing excitations is governed by a sinc function envelope modulated by a system transfer function. Accurately predicting the spall size d relies on sampling sufficient sections of this global shape to resolve mathematical ambiguities and capture the invariant amplitude relationships across varying speeds.

When the selected triplet of machine operating speeds features a large S_{span} (e.g., 300, 750, and 1500 RPM),

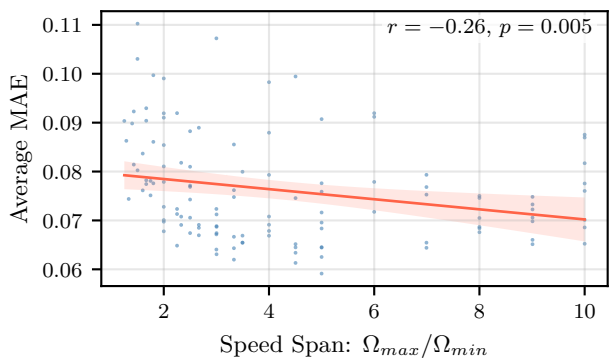

Figure 8. Mean Absolute Error (MAE) across testing speeds versus the Speed Ratio Span (S_{span}).

the neural network observes harmonic excitations spread across a wider portion of the sinc envelope. This allows the model to sample more broadly across the synthetic transfer function, potentially learning to partially compensate for its distorting effects. This expanded visibility constrains the inverse problem, granting the network the broad contextual structure necessary to output robust severity estimations that generalize across unobserved transmission paths.

Conversely, narrow speed bands with a small S_{span} (such as tightly clustered sets like 300, 332, and 375 RPM) limit the model's physical field of view regarding the underlying transfer function. By restricting visibility to a localized fragment of the combined sinc-transfer function response, the framework essentially over-optimizes via narrow curve-fitting. An example of this limitation is observed with the 1000, 1500, and 3000 RPM test combination. Despite sampling a wide absolute frequency range, the spread strictly isolates a high-frequency margin while entirely missing baseline low-frequency interactions, resulting in poor regression extrapolation when tested on speeds outside those bounds.

Robustness and Deployment Constraints. Several additional factors are expected to influence the robustness of the proposed framework under realistic deployment conditions. Measurement noise in the acquired vibration signal perturbs the harmonic amplitude estimates $c_k(\Omega_i)$ used as model inputs; while the harmonic masking regularizer applied during training provides partial robustness to corrupted or missing harmonics, the sensitivity of the framework to continuous broadband noise has not been explicitly characterized and warrants dedicated investigation. Sensor placement and mounting configuration directly shape the structural transfer function H, and the augmentation strategy, parameterized by simplified second-order resonance models, may not adequately represent extreme structural responses arising from non-standard installation positions. Load variations pose an additional challenge: the RMS normalization baseline is computed from a fixed training distribution at a single load level, and significant deviations from the training-time load regime could shift the normalized feature values outside the learned distribution, degrading prediction accuracy. These sensitivities motivate the inclusion of multi-load training data and explicit load-dependent normalization strategies as directions for future work.

From a deployment perspective, the RMS normalization procedure (Eq. 4) requires a historical fault-progression dataset from the target bearing or bearing type, meaning the current implementation is not directly applicable to a new installation from cold start. Practical deployment would therefore require adaptation of the normalization step, for example computing a baseline from an initial period of healthy operation.

Scope and Applicability Boundaries. The present framework is formulated exclusively for outer-race spall defects and relies on the sinc-based periodic pulse model for BPFO harmonic generation. Extension to inner-race defects is non-trivial: inner-race faults introduce a load-zone amplitude modulation that creates a speed-dependent envelope on the fault harmonics. Cage and rolling element defects exhibit different characteristic frequencies and generally weaker, less periodic excitation signatures, further limiting direct applicability. In the presence of multiple simultaneous faults, spectral overlap between fault harmonics would corrupt the amplitude features extracted at the BPFO lines, potentially leading to unreliable size estimates. The current framework implicitly assumes that the BPFO harmonics are cleanly separable from all other spectral components, an assumption that may not hold in multi-component or heavily degraded machine assemblies.

The bounds selected for the uniform distributions used for the transfer function generation were selected heuristically to simulate a representative range of typical structural resonances encountered in industrial machinery. The choice of one or two resonance peaks is a deliberate simplification. It is explicitly acknowledged that real industrial transfer functions are considerably more complex, potentially featuring many overlapping resonances, non-minimum phase behavior, and frequency-dependent damping characteristics. This synthetic model is therefore positioned as a controlled, proof-of-concept approximation designed to test whether the network can learn invariant harmonic ratios under moderate transfer function distortion, and not as a fully representative surrogate for real structural dynamics.

Another limitation of this formulation is that for spalls

where $d/L > 0.5$, the wrapped target $\widetilde{d/L}$ becomes identical to its mirror counterpart below the threshold, rendering these two physically distinct regimes indistinguishable to the model. In applications where severe late-stage spalling is anticipated, an additional disambiguation mechanism would be required. From a Remaining Useful Life perspective, this is a non-trivial limitation, as accurate severity estimation in the critical late-stage fault region is precisely where sizing accuracy most influences RUL predictions.

These results are obtained through internal cross-validation within a single run-to-failure experiment on one bearing specimen: the training and testing sets differ in speed combination and in the realization of the synthetic transfer functions, but are drawn from the same physical progression sequence. This evaluation therefore characterizes the ability of the model to transfer across speed combinations and synthetic structural dynamics, rather than true out-of-distribution generalization to unseen physical machines or bearing types. The reported MAE values should be interpreted with this scope in mind, and independent validation on separate run-to-failure datasets would be necessary to substantiate broader generalization claims.

6. Conclusion

In conclusion, the primary contributions of this work are:

1. **Proof-of-Concept Generalizability**: Augmenting the training data with diverse, synthetic transfer functions suggests that the model may learn to reduce sensitivity to sensor-path structural resonances, indicating a potential pathway toward application to unobserved machinery, pending validation on additional experimental setups and bearing types beyond the N209 ECP bearing studied here.
2. **Decoupled Speed Constraints**: By reducing the reliance on strict integer speed ratios, this approach shows promise as a proof-of-concept diagnostic strategy for continuous machinery, as demonstrated on a single run-to-failure experiment using any three concurrent operating speed measurements.
3. **Instantaneous Prediction**: The direct mapping strategy yields instantaneous severity estimation.

Future research will focus the validation of the methodology on independent run-to-failure datasets from diverse industrial applications which is necessary to characterize the limits of the transfer function augmentation strategy under more complex structural dynamics and multi-component environments. Second, the current model assumes that the baseline excitation signal follows the outer-race sinc-based pulse model uniformly across all machines; extending the framework to inner-race defects and multi-fault scenarios will require modifications to the kinematic feature model, incorporating load-zone modulation compensation and additional preprocessing stages. Third, the development of deployment-ready adaptive normalization strategies working with limited healthy-operation data rather than requiring the full fault progression history represents a key milestone for the deployment of as a condition monitoring system.

Acknowledgment

This work was conducted as part of the project "Smart, Aware, Integrated Wind Farm Control Interacting with Digital Twins (ICONIC)" and has received funding from the European Union's Horizon Europe Research and Innovation Actions under the grant agreement No 101122329.

References

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., & Shah, R. (1993). Signature verification using a" siamese" time delay neural network. *Advances in neural information processing systems*, 737–744.

Bublil, T., Taal, C., Maljaars, B., Klein, R., & Bortman, J. (2024). Labeling algorithm for outer-race faults in bearings based on load signal. *PHM Society European Conference*, *8*(1), 7–7.

Bublil, T., Taal, C., Maljaars, B., Matania, O., Klein, R., & Bortman, J. (2025). Experimental study on condition indicators for severity estimation of growing spalls in bearings. *Mechanical Systems and Signal Processing*, *236*, 113019.

Chen, Y., Liu, X., Rao, M., Qin, Y., Wang, Z., & Ji, Y. (2025). Explicit speed-integrated lstm network for non-stationary gearbox vibration representation and fault detection under varying speed conditions. *Reliability Engineering and System Safety*, *254*, 110596. doi: 10.1016/j.ress.2024.110596

Chopra, S., Hadsell, R., & LeCun, Y. (2005). Learning a similarity metric discriminatively, with application to face verification. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, *1*, 539–546.

Dong, Y., Li, Y., Zheng, H., Wang, R., & Xu, M. (2022). A new dynamic model and transfer learning based intelligent fault diagnosis framework for rolling element bearings race faults: Solving the small sample problem. *ISA Transactions*, *121*, 327-348. doi: 10.1016/j.isatra.2021.03.042

Fan, C., Wang, P., Zhang, Y., Ma, H., Li, X., & Wang, Q. (2025). Digital twin assisted degradation assessment of bearing cage performance. *IEEE Transactions on Industrial Informatics*, *21*(7), 5171-5181. doi: 10.1109/TII.2025.3552655

Koch, G., Zemel, R., & Salakhutdinov, R. (2015). Siamese neural networks for one-shot image recognition. *ICML deep learning workshop*, *2*.

Kotzalas, M. N., & Harris, T. A. (2001). Fatigue failure progression in ball bearings. *J. Trib.*, *123*(2), 238–242.

Liefstingh, M., Taal, C., Restrepo, S. E., & Azarfar, A.

(2021). Interpretation of deep learning models in bearing fault diagnosis. *Annual Conference of the PHM Society*, *13*(1).
Mason, J. K., Trivedi, H. K., & Rosado, L. (2017). Spall propagation characteristics of refurbished vim–var aisi m50 angular contact bearings. *Journal of Failure Analysis and Prevention*, *17*(3), 426–439.
Ngo-The, D., Taal, C., & Antoni, J. (2025). Bearing spall size estimation under varying speed conditions. *Annual Conference of the PHM Society*, *17*(1). doi: 10.36001/phmconf.2025.v17i1.4354
Ni, Q., Ji, J., Halkon, B., Feng, K., & Nandi, A. K. (2023). Physics-informed residual network (piresnet) for rolling element bearing fault diagnostics. *Mechanical Systems and Signal Processing*, *200*, 110544. doi: 10.1016/j.ymssp.2023.110544
Rosado, L., Forster, N. H., Thompson, K. L., & Cooke, J. W. (2009). Rolling contact fatigue life and spall propagation of aisi m50, m50nil, and aisi 52100, part i: Experimental results. *Tribology Transactions*, *53*(1), 29–41.
Shen, S., Lu, H., Sadoughi, M., Hu, C., Nemani, V., Thelen, A., ... Kenny, S. (2021). A physics-informed deep learning approach for bearing fault detection. *Engineering Applications of Artificial Intelligence*, *103*, 104295. doi: 10.1016/j.engappai.2021.104295
Song, Y., & Liu, P. (2024). Federated domain generalization for intelligent fault diagnosis based on pseudo-siamese network and robust global model aggregation. *International Journal of Machine Learning and Cybernetics*, *15*(2), 685–696. doi: 10.1007/s13042-023-01934-2
Vu, M.-H., Nguyen, V.-Q., Tran, T.-T., Pham, V.-T., & Lo, M.-T. (2024). Few-shot bearing fault diagnosis via ensembling transformer-based model with mahalanobis distance metric learning from multiscale features. *IEEE Transactions on Instrumentation and Measurement*, *73*, 1-18. doi: 10.1109/TIM.2024.3381270
Xu, M., Zhang, H., Miao, H., Hao, J., Li, C., Song, W., ... Zhang, Y. (2023). Model-based vibration response analysis and experimental verification of lathe spindle-housing-belt system with rubbing. *Mechanical Systems and Signal Processing*, *186*, 109841. doi: 10.1016/j.ymssp.2022.109841
Yang, X., Yang, J., Jin, Y., & Liu, Z. (2024). A new method for bearing fault diagnosis across machines based on envelope spectrum and conditional metric learning. *Sensors*, *24*(9). doi: 10.3390/s24092674
Zhang, H., Borghesani, P., Smith, W. A., Randall, R. B., Shahriar, M. R., & Peng, Z. (2021). Tracking the natural evolution of bearing spall size using cyclic natural frequency perturbations in vibration signals. *Mechanical Systems and Signal Processing*, *151*, 107376. doi: 10.1016/j.ymssp.2020.107376

Biographies

Stephan Baggeröhr is a researcher at SKF while simultaneously pursuing a Ph.D. at the Department of Mechanical Engineering, KU Leuven (Belgium). His work sits at the intersection of industry and academia, focusing on advancing bearing diagnostics and the broader implementation of applied deep learning for industrial reliability. Building on a Master's degree from the University of Pretoria, where he specialized in generative models for Prognostics and Health Management, he has maintained a career-long focus on bridging the gap between theoretical AI and practical industrial application. By maintaining his role at SKF alongside his doctoral research, he ensures his work remains grounded in the complex, real-world challenges of modern engineering.

Cees Taal is an experienced researcher in the field of sensor signal processing and machine learning. He has worked in academia—at Delft University of Technology (Delft, The Netherlands) and KTH Royal Institute of Technology (Stockholm, Sweden)—on audio and speech processing. Following his academic career, he held various industrial positions in biomedical engineering at Philips Research (Eindhoven, The Netherlands) and in the energy sector at Eneco (Rotterdam, The Netherlands). He is currently appointed as a Senior Technologist at SKF, The Netherlands, where he is responsible for defining the research strategy in the field of bearing diagnostics and prognostics. Cees received the IEEE Signal Processing Society Best Paper Award in 2016.

Konstantinos Gryllias received his Diploma and Ph.D. degrees in Mechanical Engineering from National Technical University of Athens, Athens, Greece, in 2004 and 2010, respectively. He is currently Professor of Vibro-acoustics of machines and transportation systems at the Department of Mechanical Engineering, KU Leuven, Leuven, Belgium. He also serves as the Manager of the University Core Lab Flanders Make@KU Leuven Motion Products, Belgium. His research interests include condition monitoring, signal processing, prognostics, and health management of mechanical and mechatronic systems.