CONTRIBUTION TO ESTIMATE THE LEVEL OF BEARING DEGRADATION USING AN HMM-MULTIBRANCH APPROACH

WARDHANA Indrawata
INSA CVL
Bourges
indrawata.wardhana

MTIBAA Amal Gouiaa
Orleans University
Orleans
amal.gouiaa-mtibaa@univ-orleans.fr

VRIGNAT Pascal
Orleans University
Orleans
pascal.vrignat@univ-orleans.fr

KRATZ Frédéric
INSA CVL
Bourges
frederic.kratz@insa-cvl.fr

Abstract — The degradation of industrial systems is a natural and often inevitable process. Among the methods employed to estimate this degradation, the Hidden Markov Model (HMM) stands out for its wide application. This paper focuses on maintenance strategies for rotating machines, introducing a comprehensive framework for preprocessing methods alongside a novel adaptation of the HMM known as the Extended Multi-Branch HMM (EMB-HMM). To illustrate its efficacy, the FEMTO bearing dataset was specifically selected. Initially, abnormal signals, often identifiable by their pronounced noise in frequency zones, are pinpointed. Subsequent preprocessing steps are then executed. Moving forward, the EMB-HMM framework, distinguished by its four branches and five hidden states per branch, is applied. The determination of the active branch relies on both prior and posterior probabilities, with these probabilities and branch topologies linked to the four fault frequencies. Finally, the EMB-HMM serves as the assessment model, facilitating the evaluation of bearing performance degradation and the detection of the First Predicting Time (FPT) of initial degradation.

Keywords — Prognosis and Health Management, condition assessment, data preprocessing, first predicting time, Hidden Markov Model, bearing.

I. INTRODUCTION

Rotating machinery is integral to numerous industries. The smooth operation of these machines heavily relies on the effectiveness and reliability of their bearings. Bearings serve a pivotal function in minimizing rotational friction and supporting radial and axial loads. However, due to the challenging conditions in operation, such as vibrations, torque, speed, and acceleration, bearings are susceptible to damage and potential failure. Consequently, the need for early fault detection and robust maintenance practices becomes paramount to prevent equipment malfunction, financial losses, and operational disruptions. Advanced analytical techniques, combined with a comprehensive preprocessing framework, offer a promising approach to enhance maintenance practices, improve system performance, and mitigate risks associated with bearing failure.

Detecting defects in rolling bearings at their early stages presents a significant challenge. Various types of failures, including damage to outer and inner races, ball issues, integrated malfunctions, and excessive loads, need to be accurately identified for effective maintenance (Dybała & Zimroz, 2014). Moreover, the presence of diverse technical configurations of bearings further complicates the fault detection process. Overcoming these challenges necessitates the development and implementation of robust diagnostic techniques capable of detecting faults before they develop into catastrophic damage.

Vibroacoustic analysis emerges as a widely employed method to detect technical degradation in rolling bearings at early stages. This technique involves studying the intricate vibration signals produced by the interaction of various components, such as the outer race, rolling elements, inner race, and cage, in a rolling bearing. By analyzing the modulation of these signals at characteristic frequencies, induced by mechanical impacts resulting from defects, valuable insights regarding defect localization and severity can be obtained. Vibroacoustic analysis enables the timely identification of potential faults and facilitates proactive maintenance practices.

Prognostics and Health Management (PHM) plays a crucial role in smart maintenance methods (Lei *et al.*, 2018). PHM strategies aim not only to detect faults but also to propose a time window during which the system can continue to operate, even in a degraded state. This comprehensive approach to maintenance encompasses various steps, including data acquisition and preprocessing, where data quality verification is essential for reliable diagnosis and prognosis. Condition assessment, prediction of FPT, and diagnostic and prognostic steps, such as estimating Remaining Useful Life (RUL), contribute to effective maintenance decision-making and improved system reliability.

A method for extracting analysis signals is essential for providing relevant insights into the health condition of the bearing. Additionally, noise and vibrations stemming from other system parts can induce irregular fluctuations. If left unaddressed, these fluctuations could trigger false alarms and compromise the accuracy of failure prognosis and prediction, as noted by Chen (K. Chen *et al.*, (2023)). In the early phases of PHM, assessing the current system condition is often the first step toward predicting RUL. The forecast of RUL relies heavily on FPT, which is intricately connected to the health condition of the bearing, as depicted in Figure 1.

Sophisticated diagnostic and prognostic methods are employed to analyze data and make informed decisions about bearing health. Data-driven models, including Statistical Learning (SL) and Machine Learning (ML) (Vrignat *et al.*, 2022), offer valuable insights into the detection and prediction of bearing degradation. One notable approach within the realm of data-driven models is the HMM, which combines elements of both ML and SL.

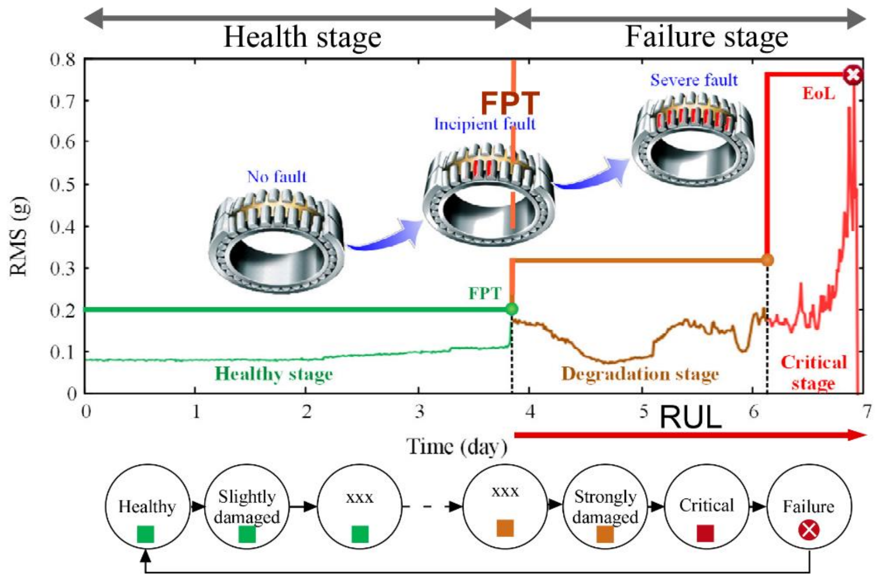

Figure 1. Bearing degradation states with FPT and Stages (adapted from Lei *et al.,* (2018b); Vrignat *et al.,* (2019); Zhu *et al.,* 2022).

Maintaining the robustness and reliability of rotating machinery, particularly through effective bearing maintenance, is of utmost importance across diverse industries. Advanced analytical techniques, such as vibroacoustic analysis, and the implementation of PHM strategies enhance maintenance practices by enabling early fault detection and timely decision-making. Diagnostic and prognostic methods, including the employment of data-driven models like Multi-branch HMM, contribute to precise assessment of bearing health and accurate estimation of RUL. By leveraging these approaches, industries can improve system performance, mitigate risks associated with bearing failure, and optimize maintenance practices for overall operational efficiency.

The subsequent sections of this paper follow a structured outline. Section 2 provides a literature review about the theories of preprocessing and HMM. Section 3 outlines the framework of the proposed method. In Section 4, the effectiveness of the proposed method is demonstrated through experiments using FEMTO datasets. Finally, Section 5 provides a summary of the findings and derives conclusions from the outcomes of the study.

II. STATE OF THE ART

Rolling element bearings (REB) are essential components in rotating machinery, serving the purpose of minimizing rotational friction caused by heavy loads or ineffective sealing. However, the failure of inner races, ball issues, rolling elements, etc can lead to catastrophic machine breakdowns (Dybała & Zimroz, 2014). Detecting defects at their early stages is a key challenge in rolling bearing diagnostics, as it allows for timely alerts to machine operators, preventing severe damage. Early-stage fault diagnosis methods are crucial for maintaining the robustness of rolling bearings, as the degradation process is continuous and can last for a significant period of time. Ensuring the reliability and performance of REB requires a proactive approach to defect detection, ultimately enhancing overall machine efficiency and preventing costly breakdowns *(Mohanty et al.,* 2018)..

REB are complex vibration systems, where the interactions between their components give rise to intricate vibration signals. These signals are modulated by mechanical impacts that occur when faults on the bearing surfaces strike each other. By analyzing these vibration signals using vibroacoustic analysis, Mohanty *et al.,* (2018) have been able to detect most of the damages in the earlier stages of bearing degradation. The characteristic frequencies of the bearing faults can be categorized into different zones, such as the bearing defect frequency zone, shaft speed zone, bearing natural resonances zone, and high-frequency zone, as observed by Eren (2017). These frequency zones play a crucial role in identifying and localizing defects within the bearing (Xie et *al.,* 2013). Figure 2 provides a visual representation of the frequency content for all four stages of bearing degradation, giving insights into the progression of damage and the associated characteristic frequencies.

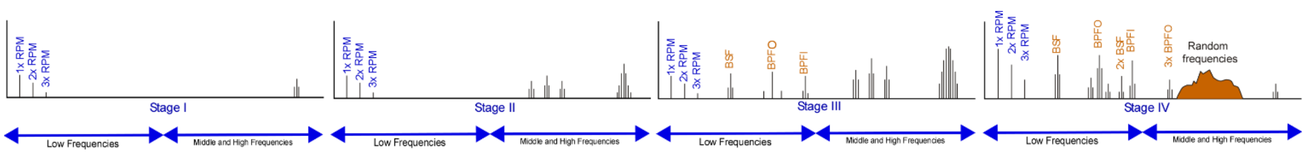

Figure 2. Four stage bearing fault and multiple zone (adapted from Eren (2017)).

Our objective is to estimate the bearing degradation level. To achieve our challenge, we need to do the link from data acquisition to prognostic. Preprocessing steps play a pivotal role in enhancing the accuracy and effectiveness of the prognostic process. These steps involve noise reduction, normalization, outlier detection, feature selection, feature reduction, missing values, and feature extraction (Amasyali & El-Gohary, 2018; Liu *et al.*, 2021; Khan *et al.*, 2022). The Signal-to-Noise Ratio (SNR) is the power ratio between the part of the signal that represents information and the rest, which constitutes background noise. It is an indicator of the quality of information transmission. Low SNR related to high noise of vibration. In order to gain a high SNR (C. Liu *et al.*, 2022), the primary focus of investigation has been on exploring the separation of low-frequency signals that convey information about bearing rotation and faults from the middle and high-frequency components. The Savitzky-Golay Filter (SGF) stands out as a robust tool for denoising and effectively smoothing signals originating from faults (Biswal *et al.*, 2022).

An analysis signal extraction method is necessary to offer pertinent insights into the bearing state of health. In fact, noise and vibrations from other system components can also lead to abnormal fluctuations as shown in Figure 3 (a). If these fluctuations are not dealt with, they may cause false alarms and affect the judgment of FPT and the prediction accuracy (Chen et al., 2023). Therefore, random anomalies and spurious fluctuations in Health Index (HI) can greatly affect the state estimation. In health monitoring, the objective is to determine the FPT, which serves to partition the operational phases of the bearing into health and degradation stages, as illustrated in Figure 1. L. Wang *et al.*, (2022) used fixed abnormal threshold for defining FPT, while Ding et al. (2022) use tolerance parameter as threshold. Morevoer, FPT is identified when the faulty states manifest consecutively (J. Zhu et al., 2020). In prediction process, the FPT also used as classification accuracy of bearing degradation. In Figure 3 (b), it is evident that there are instances of misclassification of states, referred to as Fake Normal (FN) and Fake Fault (FF).

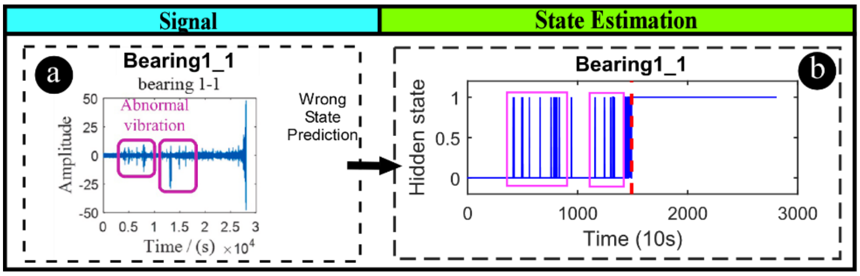

Figure 3. (a) Abnormal signal vibration (adapted from (H. Wang et al., 2022)) (b) Error diagnosis related to FF and FN(adapted from (J. Zhu et al., 2020)).

HMM stands as one of the widely acknowledged techniques for gauging the extent of degradation within a system (Peng & Dong, 2011). These models leverage hidden states inferred from observation bearing reading. Due to its proficiency in handling sequential data, it enables the analysis of changing conditions over time. The inherent probabilistic framework of HMM is adept at handling uncertainties arising from sensor noise, incomplete data, and fluctuations in system conditions, thereby enhancing the reliability and robustness of assessments. As illustrated in Figure 3, erroneous sensor readings result in imprecise state predictions. Additionally, it is noteworthy that multiple deteriorations in bearings may occur concurrently, a scenario not effectively addressed by conventional HMM approaches, which typically predict only a single deterioration at a time. Le et al. (2016) solve this problem with introduced a research on applying HMM in multiple deteriorations using fatigue cracks datasets. This approach involves HMM and Hidden Semi-Markov Models (HSMM) with a left-to-right automaton topology (Le *et al.*, 2015; Le *et al.*, 2016). In this reseach, we tried to use MB-HMM in real industrial application, actually from bearing dataset measurement using our topology probabilities.

III. METHODOLOGY

A. *Case Study*

The FEMTO institute supplied a dataset consisting of 17 instances of rolling element bearing failures. These failures were monitored through the vibration signals and temperatures recorded using two accelerometers and a thermocouple. Termination of a bearing's useful life was identified when the vibration signal amplitude exceeded 20 g across various operational conditions (Lei et al., 2018). The dataset lacks specific information regarding the location of bearing defects, such as outer defects. Thus, our analysis focuses on a specific operational condition (OC) 1, characterized by a load of 4000N and a shaft speed of 1800 rpm. The training set comprises data from bearing1-1 and bearing 1-2, while the testing set encompasses data from bearing1-3, 1-4, 1-5, 1-6, and 1-7. The dataset was selected due to its unique characteristics, characterized by unknown origin deteriorations.

B. *Proposed Bearing Degradation Approach*

The main target in this study is find the degradation states of the bearing and finding the FPT using a HMM. For investigating the states degradation from HI, we should use two main steps. The first steps is producing high quality HI using preprocessing methods. Then, second steps is to find the states condition using HMM.

1. Step 1

Preprocessing in bearing analysis involves essential steps to prepare raw data for accurate fault identification. Feature extraction techniques focus on identifying critical characteristics in sensor data, including frequency, time-domain, and other relevant features. In bearing condition monitoring, detecting periodicity or recurring patterns in vibration signals is crucial for early problem detection. Maximal Overlap Discrete Wavelet Transforms (MODWT) offer a robust method for analyzing machine data with significant periodicity, maintaining detail across all decomposition levels. MODWT decomposes the initial signal into distinct components: a low-pass filtered approximation A_j and high-pass filtered detail components D_j, W_j and V_j represent the wavelet and scaling coefficients, low-pass filter $\tilde{g}_j$ and the high-pass filter $\tilde{h}_j$ at the j^{th} level. The mathematical framework for MODWT is defined by Equations (1) to (3), detailed by Seo *et al.*, (2017).

$$X = \sum_{j=1}^{L} D_j + A_{J0} \quad (1)$$

$$D_{j,t} = \sum_{l=0}^{n-1} \tilde{h}^{0}_{j,l} W_{j,t+l \text{ mod n}} \quad (2)$$

$$A_{j,t} = \sum_{l=0}^{n-1} \tilde{g}^{0}_{j,l} V_{j,t+l \text{ mod } n} \quad (3)$$

After refining the signal, we employ the HI method to detect degradation. We utilize time-domain features for this purpose. Time-domain features are commonly employed for feature extraction in bearing analysis, including Mean, Standard variance, RMS , Square of RMS, Skewness, Kurtosis, Crest Factor, Margin Factor, Form Factor, Impulse Factor, Maximum, Minimum, Peak-to-Peak value, Absolute Mean, Absolute CF, Energy (Y. Liu *et al.*, 2021). To identify anomalies, we recommend for employing the z-score method, which gauges deviations from the mean in terms of standard deviations. While common thresholds include values like 2.5, 3, and 3.5 (Nkikabahizi *et al.*, 2022), these can vary based on specific contexts. The Savitzky-Golay filter (SGF) proves instrumental in checking vibration signals from bearings. Unlike conventional moving average filters, SGF implements a least-squares polynomial fitting within its moving window, enabling preservation of crucial signal characteristics such as peak heights and waveform shapes. This preservation is pivotal for precise fault detection (Atif & Khalid, 2020). To find the best features from time domain and to comprehensively assess the degradation process of rolling bearings of time domain features, we use feature selection trend consistency effectively captures the underlying characteristics (Lei et al., 2018). All preprocessing methods are describe more detail based on the flow in Wardhana *et al.*, (2023).

$$\text{Cor}(X) = 1 - \frac{6 * \sum_{k=1}^{n} \left(X_T(t_{k+1}) - X_T(t_k) \right)}{n(n^2-1)} \quad (4)$$

2. Step 2

In this stage, we will employ various methods to delineate the degradation states. The methodologies for each method are delineated in Framework Figure 4. HMM are statistical constructs utilized for depicting stochastic processes. Unlike conventional Markov chains, in an HMM, the states are not readily observable. A comprehensive depiction of an HMM encompasses:

- N: number of states of the Markov chain. States are denoted as $Q = \{Q_1, \dots, Q_N\}$, where the state at time t is denoted as q_t
- M: number of observations $O = \{O_1, \dots, O_T\}$, where $O_t \in V$ is the observation at time t with $V = \{v_1, \dots, v_m\}$ is the set of observable values.
- *π:* initial state probability $\pi = \{\pi_i\}$; with $\pi_i = \text{P}(q_1 = i), 1 \le i \le N$
- *A*: state transition probability matrix $A = \{a_{ij}\}$; with $a_{ij} = \text{P}(q_{t+1} = j \mid q_t = i), 1 \le i, j \le N, \sum_{ij} a_{ij} = 1$
- *B*: emission probability distribution $B = \{b_j(k)\}$, where $b_j(k) = \text{P}(O_t = v_k \mid q_t = Q_i)$ $,1 \le j \le N, 1 \le k \le M$

Then, an HMM models can be describing as compact notation $\lambda = \{\pi, A, B\}$. In practice, HMM are used to solved the following typical problems (Rabiner, 1989):

- Problem 1: The Forward Backward algorithm used to compute the probability $\text{P}(O / \lambda)$ of being in a particular state at a certain time, given all the observed data $O = \{O_1, \dots, O_T\}$. This problem known as evaluation problem (in prognostic is consider as a diagnostic problem) (Tobon-Mejia *et al.*, 2010).
- Problem 2: Given some O and some λ, find the best state sequence $Q = \{Q_1, \dots, Q_T\}$ given the observed data. The Viterbi algorithm solves this problem (Viterbi, 1967).
- Problem 3: find the model parameter $\lambda = (\pi, A, B)$ that better fit the observation sequence O. The Baum-Welch algorithm used to estimate the parameters of an HMM $\lambda^* = \arg\max_{\lambda} \text{P}(O / \lambda)$ from observed data (Baum, 1972).

Selecting the topology in HMMs involves making choices about the ideal model setup, which includes determining the appropriate number of hidden states and transition probabilities. The goal of this selection process is to identify the most fitting model that accurately reflects the patterns present in the data. This process including selecting the HMM type and specifying the number of states to be used. HMM topology have several types: Ergodic HMM (Zimmerman *et al.*, 2018), Linear HMM (Figueroa-Angulo *et al.*, 2015), Left-to-right HMM, Cyclic left-to-right HMM (Prakash *et al.*, 2017). HMM are typically employed with discrete data. The problem with this approcah is that the accelerometer observations are continous signal. To convert continuous data into discrete form, clustering methods are utilized. Gaussian Mixture Models (GMM) are recognized as effective clustering methods compared to k-means. In the optimization of GMMs, the Expectation-Maximization (EM) algorithm is employed to minimize the negative log-likelihood, thereby facilitating the learning process (Chaleshtori & Aghaie, 2024). In Figure 5(a), the flowchart explain the topology selection and clustering.

By experimenting with varying numbers of states within the HMM, the model's performance can be assessed and refined to attain the optimal fit to the data. In Figure 5, we present several topologies tailored to address industrial challenges. These topologies are associated with fault frequencies stemming from bearing deterioration. The sequence begins with state starting from S_1 and concludes with failure state denoted as S_5. These four topologies correspond to four branches. To identify the active branch, we establish it using prior probability, denoted as ϕ ,where $\phi = \{\phi_1, ..., \phi_p\}$ with *p* branches number.

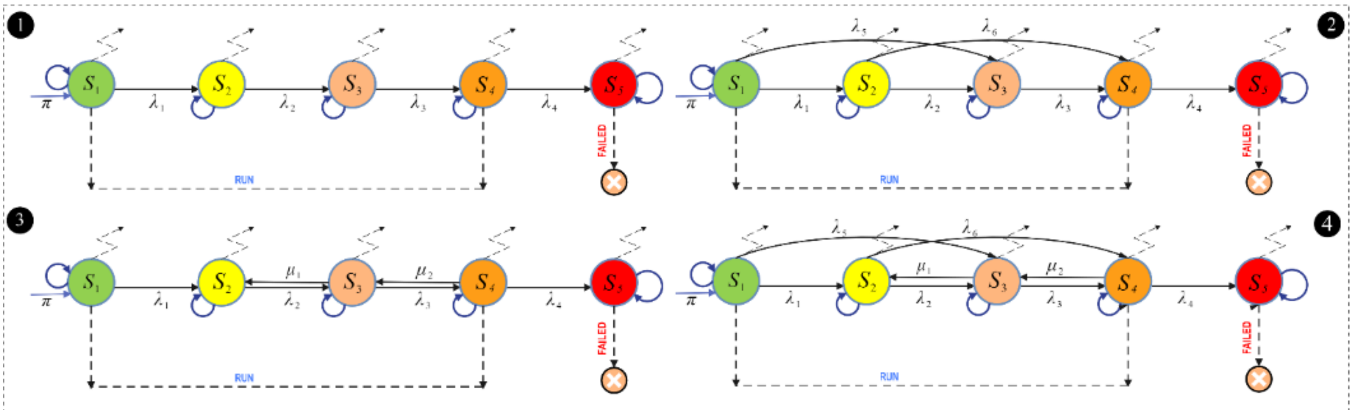

Figure 4. Four type of bearing topology proposed

Le *et al.*, (2016) define prior probability using the number of fatigue crack types. However, FEMTO bearings lack information about deterioration. Thus, we define prior probability by extracting signals into the frequency domain using FFT and relating them to the number of frequencies as the number of faults. We also use FFT values as training data. One problem with transition probability is defining the values of A. Typically, researchers use static values (Celeux & Durand, 2008), but this approach could time suffering and decrease likelihood. To improve this, we train the data using the Baum-Welch learning method to produce the optimal A for each branch described in Figure 5(b). One of transtition probability problem is to define the values of A. Commonly, researchers use static values, but the defining with could wasting time and decrease the likelihood. In order to improve that, we train the data with Baum-Welch learning method produce the optimal $\lambda^* = (\pi^*, A^*, B^*)$ for each branchs that described in Figure 5(b). To find the prior probability, we utilize equations (7) through (10) as explain in Figure 5(c).

$$S = \sum_{i=1}^{N} \sum_{j=1}^{M} \delta(x_{ij}) \quad (5)$$

$$\delta(x_{ij}) = \begin{cases} 1 & \text{if } x_{ij} = 1 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$f_{ij} = \frac{x_i}{\sum_{j=1}^{M} x_{ij}} \quad (7)$$

$$\phi_j = \mathrm{P}(\lambda_m) = \frac{\sum_{i=1}^{N} f_{ij}}{S} \quad (8)$$

S represents the total count of data points. The variable i represents time series from 1 to N, while *j* represents the number of fault frequencies from 1 to M. δ is the Kronecker delta function, which x_{ij} equals 1. f_{ij} represents the probabilities of x_{ij} . ϕ_j is the prior probability of branch M. The specific degradation mode can be identified as the one with the highest posterior probability when considering the test data sequence. Posterior probabilities of each model can be calculated via the bayes theorem (Eq.9).

$$\mathrm{P}(\lambda_m \mid O) = \frac{P(O \mid \lambda_m) P(\lambda_m)}{\sum_{m=1}^{M} P(O \mid \lambda_m) P(\lambda_m)} \quad (9)$$

$$\hat{m} = \arg\max_{m} \mathrm{P}(\lambda_m \mid O) \tag{10}$$

where $\mathrm{P}(O \mid \lambda_m)$ represents the likelihood of the model λ_m with respect to the observation sequences O, and $\hat{m}$ refers to identifying the degradation mode. After the deterioration detected, we use viterbi algorithm to define the states condition. Specifically, suppose that the equipment is currently in the mode m and let Q_m denote a possible path of states under the mode m. The optimal sequence of states is determined by:

$$Q^* = \arg\max_{Q_{\hat{m}}} \mathrm{P}(Q_{\hat{m}} \mid O, \lambda_{\hat{m}}) \tag{11}$$

The current health state of the equipment corresponds to the last state in the sequence Q^*. The diagnostic process of bearing degradation health is described in Figure 5 (b), (e), and (f). In training phase, the branches are process with training data OC1. Process of each branch equal to standard HMM process. The continuous signal transform into discrete with cluster concepts using GMM method with prior mean and covariance.

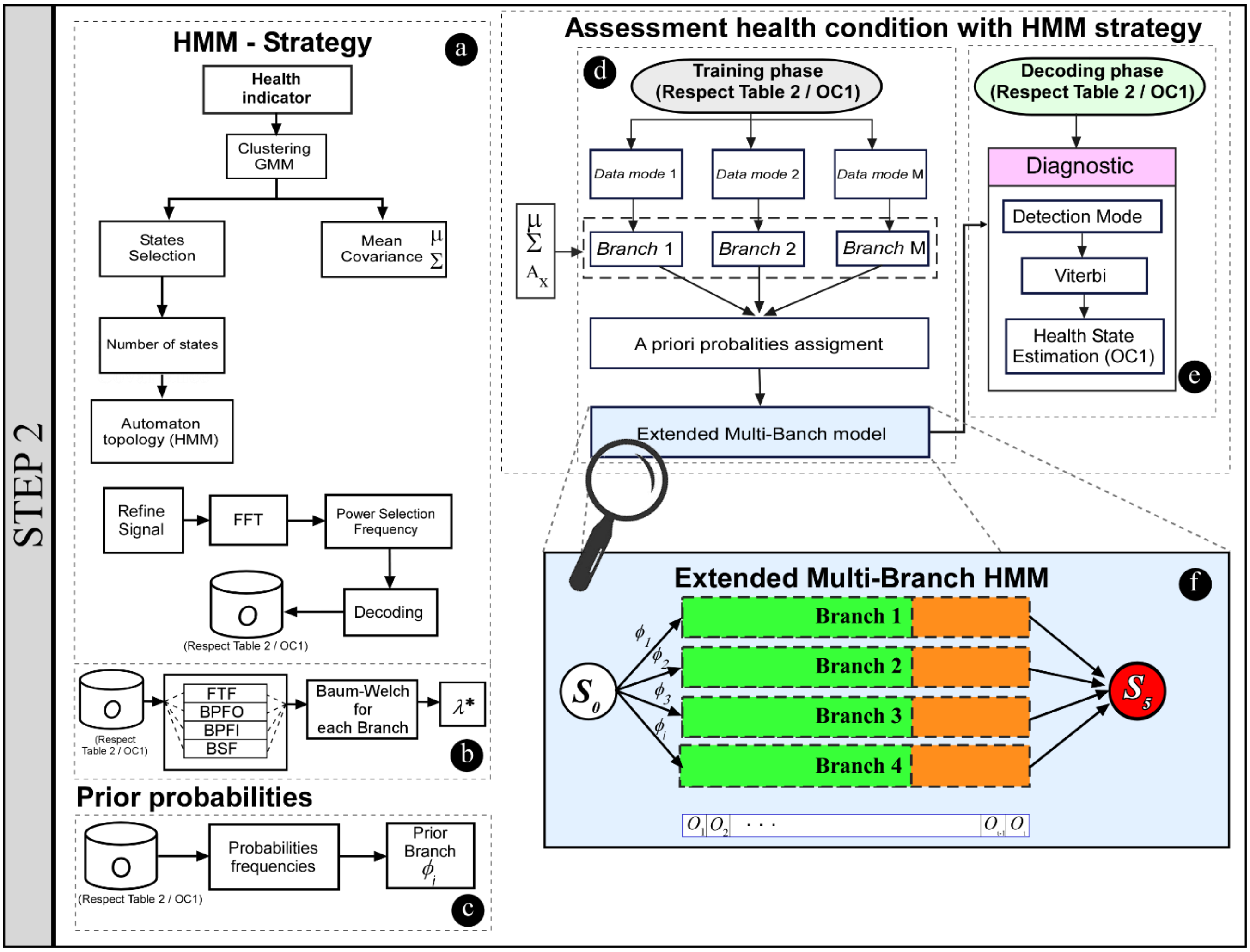

Figure 5. Proposed Framework for diagnosis degradation states.

The FPT is identified when the first faulty states occur continuously with threshold occurrence number. Let $S = [s_1, s_2, s_3, ..., s_n]$ represent the time series of tuples (t_i, ϑ_i), where t_i represents the time index and ϑ_i represents the data point of state position at time index t_i. The FPT should follow this mathematically condition:

$$\begin{cases} \vartheta_{t_{start}} = \vartheta_{t_{start+1}} = \ldots = \vartheta_{t_{start+(sx-1)}} = S_f \\ \vartheta_{t_{start-1}} \neq S_f \\ \vartheta_{t_{start+sx}} \neq S_{f-1} \end{cases} \tag{12}$$

where, t_{start} is the time index where the search for consecutive occurrences begins, $\vartheta_{t_{start}}$ is the data point at time index t_{start}, and number of threshold consecutive occurrences. In the end, we can the FPT using this equation:

$$FPT_f = \min\{i : d_i = s_f\} \quad (13)$$

where, $\min\{i : d_i = s_f\}$ represent the set of all indices i for which s_i equals s_f. The time FPT_f is the smallest index at index at which s_f is found in the data sequence S. By defining the FPT, we got two types of misclassifications (Ding *et al.*, 2022):

- Fake Fault (FF): These fault alarms occur when the system detects a fault in a normal condition before the FPT.
- Fake Normal (FN): This situation arises when the system predicts a normal condition after the FPT.

IV. RESULTS AND DISCUSSIONS

In Step 1, preprocessing methods are introduced, with deeper concern about noise and abnormal function of the signal from accelerometer bearing 1-2 as shown in Figure 7(a). In the red dashed line, it can be observed that the signal fluctuates over time. The vibration signal was extracted into multiple time-domain features (mean, Kurtosis, standard deviation, …). Standard deviation (std) showed the highest score for feature selection using the trend consistency method. Moreover, this time-domain feature contains a lot of noise in signals over time, similar to the raw signal. For example, one measurement at sampling number 70 is plotted in Figure 7(b) on the left. After data point 2000, there is a signal spike around 10 g possibly related to noise. This signal will impact the standard deviation value. For more detail, we attempted to extract the signal into the time-frequency domain over time. It is shown that all frequencies exist with high power in the frequency range of 2-10 kHz. Meanwhile, the shaft frequency, bearing fault frequencies, and harmonics exist below the frequency of 1 kHz. To address this, we use the decompose method MODWT. By utilizing the db2 wavelet, we process all signals in bearing 1-2. The decomposed signal is separated into low, medium, and high frequencies. In this process, the low frequency is chosen. The refined signal is shown in Figure 7(b) on the right. It clearly explains that the noise comes from the frequencies above 1 kHz, where the spike is reduced as shown in the red line. Additionally, in all frequency ranges, the high frequencies are reduced. Finally, the clean signal is introduced in Figure 7(c), with the standard deviation feature exhibiting a smooth line.

It is also essential to consider very high spikes during the lifetime of the bearing. These could be related to the bearing degradation moving to the next stage. Based on raw signal data, the vibration signal in bearing 1-1 around 1400 sample points exceeded 20 g, which, according to references, is known as a failure stage. Similarly, using the standard deviation feature from the refined signal, several outliers were detected in bearing 1-1 in Figure 6(a). A closer look with a zoom-in in Figure 6(b) reveals that, during the time sampling from 1300 to 1400, the surge could be handled with a noise reduction process. The SGF method was adopted to follow the trend of the signal. Moreover, the signal of the HI transform into a similar range with normalization function.

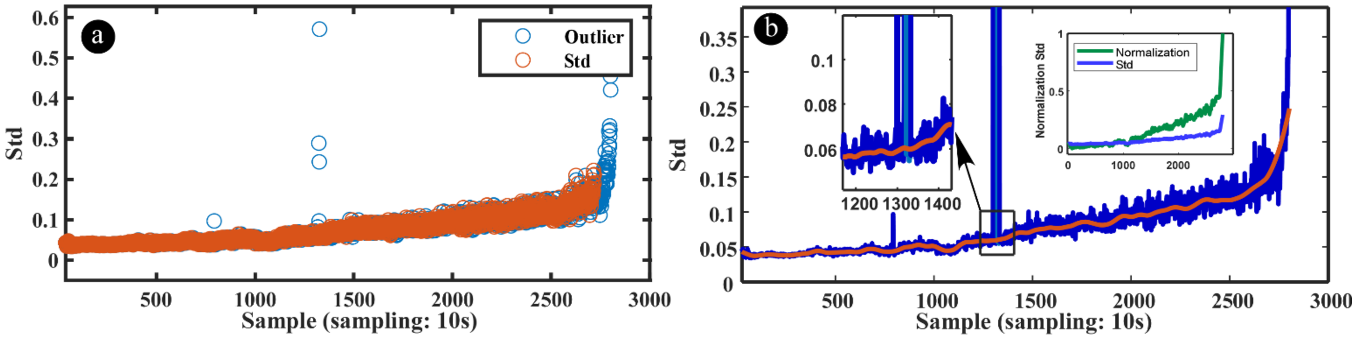

Figure 6. (a) Outlier detection (b) Noise elimination and normalization

In second step based on framework. The process of determining states and deriving statistical properties involves utilizing a combination of HMM evaluation criteria for finding optimal states and GMM clusters for producing discrete data. Based on the crack surface (Kittle, 2023), bearing have five different stage of cracks. We define it as five clusters. The clustering results can be seen in Figure 8. The clustering process, employing variables such as standard deviation (X1) and margin factor (X2), yields mean and covariance with diagonal leaning-parameter covariance. These statistical properties play a pivotal role in enhancing the accuracy of state predictions while concurrently minimizing processing time.

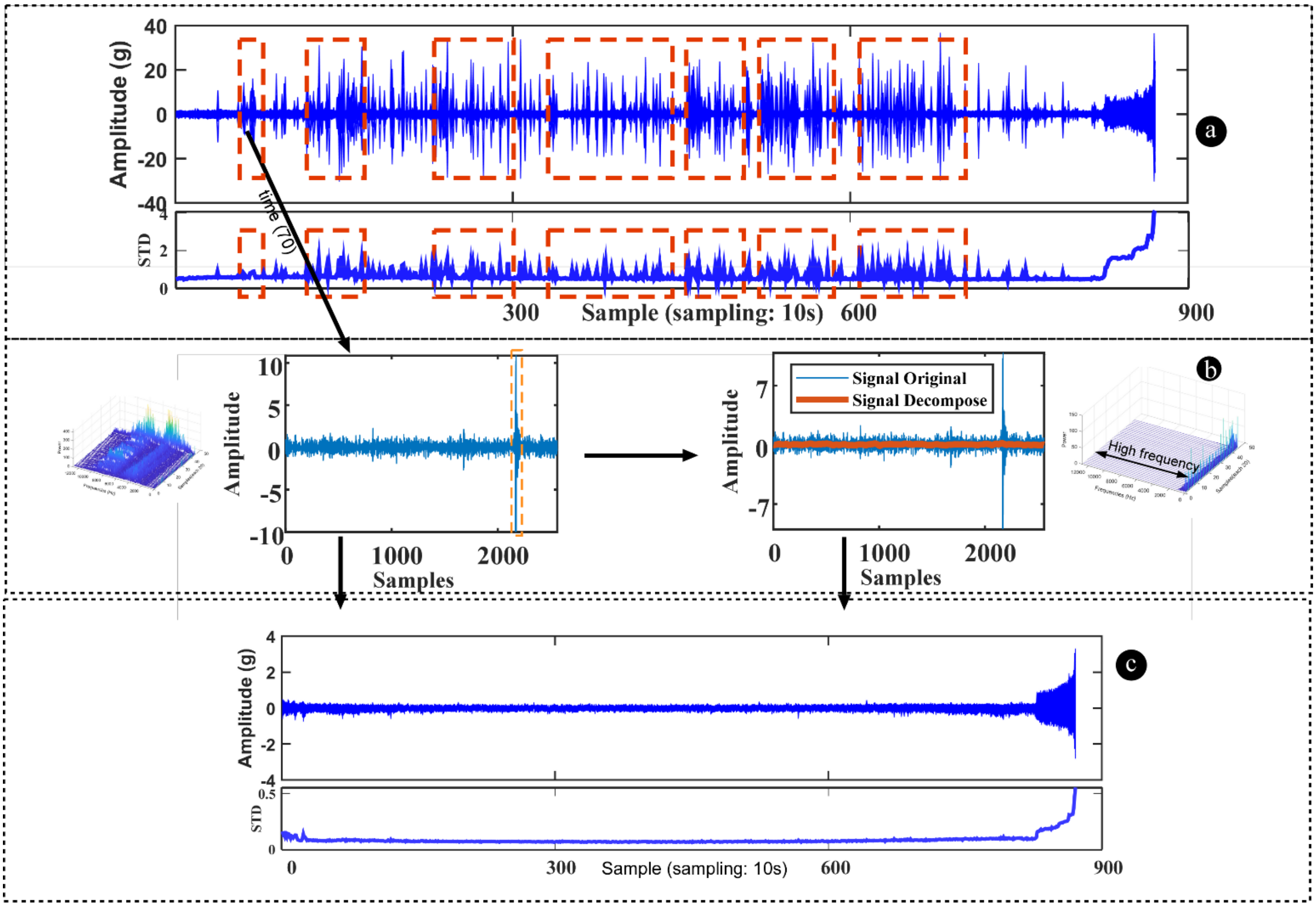

Figure 7. Decompose signal process in bearing 1-2.

We obtain the branch topology for each specific mode through training with discrete that resulted from the comparison of frequency faults as seen in Figure 9 (a). To identify the topologies linked to frequency faults, we use DHMM training with 4x4 combination topologies and discrete frequency data covering FTF, BPFO, BSF, and BPFI. Within Figure 4, these topologies are designated as Branch 1, Branch 2, Branch 3, and Branch 4. Subsequently, we found prior probabilities $\phi = \begin{bmatrix} 0.17 & 0.61 & 0.04 & 0.19 \end{bmatrix}$ for each branch. These probabilities are derived using Equation 9.

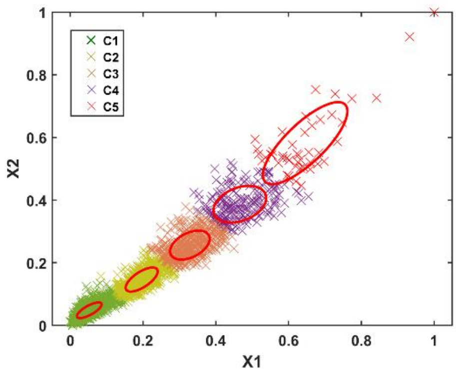

Figure 8. Clustering time-domain feature

The process of training the EMB-HMM involves combining different statistical methods. We start with parameters from the GMM, using mean and covariance values for producing cluster data. This process repeats 10 times, also considering prior probabilities. The model we create has four different structures, and it represents five states labeled S_1 to S_5. The likelihood of each branch is determined using the forward-backward algorithm to process the posterior probabilities. For an example with bearing 1-2: as time passes, we see branch 4 becoming more dominan to 1, while the others become less, as shown in Figure 9 (b). Moreover, we use the viterbi algorithm to find the best path through the states.

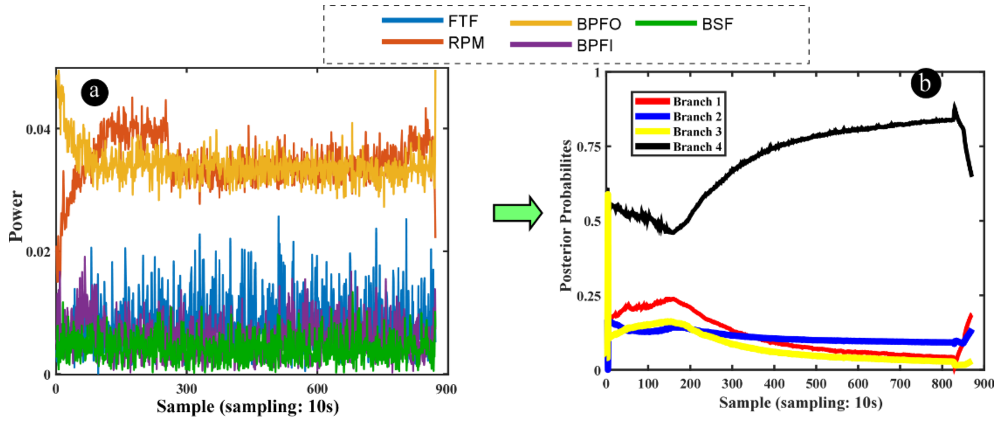

Figure 9. (a) Evaluation Criteria (b) Clustering time-domain feature

By employing the MB-HMM strategy, we can establish the active branch as the active mode for each bearing, which correlates with the deterioration of the bearing. For instance, in bearing 1-3, we identified the active branch as 4, whereas in other test bearings (1-5 to 1-7), the active branches are 2. These findings enable us to define the states of degradation. Leveraging the Viterbi algorithm with the active branch, we can determine the most likely sequence of hidden states for the bearing's condition.

In Figure 10 (a), we observe the system prediction of various states. The sampling of the bearing condition begins at time sample 0 and continues until sample 2376. Initially, the system remains in state s_1, indicating that the bearing is still in a normal condition. Subsequently, it transitions to state s_2 at time sample 1720, signaling a potential fault prediction. The shift from state s_1 to s_2 is associated with the FPT. The criterion outlined in equations 12-13 for finding FPT allows us to employ a threshold of consecutive occurrences set at 3 point. For instance, we find that the minimum time index of state s_2 is 1720, using threshold consecutive occurrences at 3, we got the position $\vartheta_{t_{1720}} = \vartheta_{t_{1721}} = \vartheta_{1721} = s_2$, $\vartheta_{t_{1719}} \neq s_2$, $\vartheta_{t_{1724}} \neq s_1$ where this criterion meet the equation 12 and 13. So, the FPT is 1720. In the case of bearing 1-3, no false faults or false normals are detected. In contrast, in sampling of about 2150, a single point returns from state s_3 to s_2, with only two points present in state S_3 and S_4. The machine continues to operate until it forecasts a failure, represented by state s_5 at the end of the observation period. Unlike, the bearing 1-3, other states bearing prediction exist more than one point. For instances, in bearing Moreover, from bearing 1-4 to bearing 1-7, no fake faults and fake normal are observed. Each of the bearings effectively identifies the FPT through the EMB-HMM method, as indicated by the red lines on the graph. To learn deeper into the specifics of state degradation, we zoom in closely to the FPT.

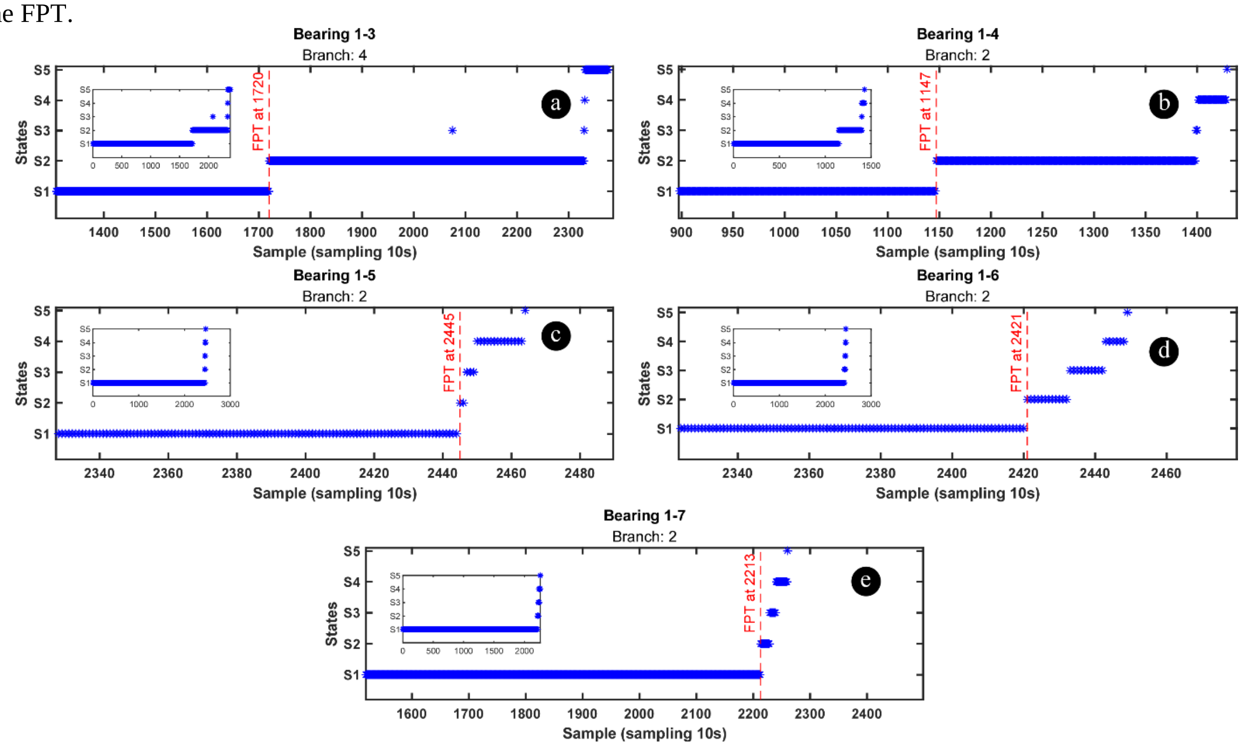

Figure 10. Degradation states of bearing 1-3 to bearing 1-7

The results of the FPT can be compared with other results from Table 1. It can be see that, the comparison of several articles give different values. The table compares the performance of different fault detection methods across various bearings. The Propose method consistently performs well, particularly for B1-5. Other methods like Auto Regressive (AR)+ Wavelet

(WT) and Long Short-Term Memory (LTSM) + Wavelet Feature Enhancement (WFE) show moderate performance across bearings, while methods like HMM and Dual Temporal Domain Adaptation (DTDA) exhibit mixed results. Self-Supervised Pre-training via Contrast Learning (SSPCL) and Adaptive Multi-scale Morphological Analysis (AMMA) + Bandwidth based Empirical Mode Decomposition (BEMD) also perform moderately, with strengths in specific scenarios.

Table 1. Comparison results from Multiple FPT Method

Method	**B1-3**	**B1-4**	**B1-5**	**B1-6**	**B1-7**
Propose	**1720**	**1147**	**2445**	**2421**	**2213**
AR+WT (Jin *et al.*, 2016)	1842	1108	1653	1656	2233
HMM (Zhu *et al.*, 2022)	1684	1083	680	649	1026
DTDA (Mao *et al.*, 2020)	1000	896	1052	1341	1175
LTSM + WFE (Li *et al.*, 2017)	1613	1082	2306	2035	2030
SSPCL (Ding *et al.*, 2022)	967	875	1047	857	956
AMMA+BMMD (Li *et al.*, 2017)	1870	1290	2240	1890	2140

V. CONCLUSIONS

In this study presents a comprehensive approach to preprocess vibration data and determine states for effective bearing fault detection. By employing meticulous signal preprocessing techniques, which encompass signal decomposition, noise reduction, and outlier handling, the refined signals manifest enhanced clarity, thus facilitating more precise feature extraction. The utilization of EMB-HMM within a systematic framework facilitates the identification of optimal states and clusters, enhancing the overall performance of fault detection systems. By evaluating model performance using metrics such as AIC and BIC, and training branch topologies, our approach ensures robustness and reliability in fault prediction. The proposed EMB-HMM in this investigation elevates the precision of state sequence estimations, diminishes false alarms, and delivers accurate estimations of state changes. Overall, this study contributes to advancing fault detection methodologies, with implications for enhancing machinery reliability and maintenance strategies.

REFERENCES

Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, *81*(September 2017), 1192–1205. https://doi.org/10.1016/j.rser.2017.04.095

Baum, L. E. (1972). An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes. In O. Shisha (Ed.), *Inequalities {III}: {P}roceedings of the Third Symposium on Inequalities* (pp. 1–8). Academic Press.

Celeux, G., & Durand, J.-B. B. (2008). Selecting hidden Markov model state number with cross-validated likelihood. *Computational Statistics*, *23*(4), 541–564. https://doi.org/10.1007/s00180-007-0097-1

Chen, K., Liu, J., Guo, W., & Wang, X. (2023). A two-stage approach based on Bayesian deep learning for predicting remaining useful life of rolling element bearings. *Computers and Electrical Engineering*, *109*(PB), 108745. https://doi.org/10.1016/j.compeleceng.2023.108745

Ding, Y., Zhuang, J., Ding, P., & Jia, M. (2022). Self-supervised pretraining via contrast learning for intelligent incipient fault detection of bearings. *Reliability Engineering and System Safety*, *218*(PA). https://doi.org/10.1016/j.ress.2021.108126

Dybała, J., & Zimroz, R. (2014). Rolling bearing diagnosing method based on empirical mode decomposition of machine vibration signal. *Applied Acoustics*, *77*, 195–203. https://doi.org/10.1016/j.apacoust.2013.09.001

Eren, L. (2017). Bearing fault detection by one-dimensional convolutional neural networks. *Mathematical Problems in Engineering*, *2017*, 1–10. https://doi.org/10.1155/2017/8617315

Jin, X., Sun, Y., Que, Z., Wang, Y., & Chow, T. (2016). Anomaly detection and fault prognosis for bearings. *IEEE Transactions on Instrumentation and Measurement*. https://doi.org/10.1109/TIM.2016.2570398

Khan, I. U., Javeid, N., Taylor, C. J., Gamage, K. A. A., & Ma, X. (2022). A Stacked Machine and Deep Learning-Based Approach for Analysing Electricity Theft in Smart Grids. *IEEE Transactions on Smart Grid*, *13*(2), 1633–1644. https://doi.org/10.1109/TSG.2021.3134018

Kittle, R. (2023). Modern Vibration Analysis Technology. In *SPM Instrument*.

Le, T. T., Bérenguer, C., Chatelain, F., Le, T. T., Bérenguer, C., Mod-, F. C. M. H. S., & Le, T. T. (2015). Multi-Branch Hidden Semi-Markov Modeling for RUL Prognosis. *2015 Annual Reliability and Maintainability Symposium (RAMS)*. https://doi.org/10.1109/RAMS.2015.7105132

Le, T. T., Chatelain, F., & Bérenguer, C. (2016). Multi-branch hidden Markov models for remaining useful life estimation of systems under multiple deterioration modes. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *230*(5), 473–484. https://doi.org/10.1177/1748006X15624584

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, *104*, 799–834.

https://doi.org/10.1016/j.ymssp.2017.11.016
Li, Y., Xu, M., Liang, X., & Huang, W. (2017). Application of bandwidth EMD and adaptive multiscale morphology analysis for incipient fault diagnosis of rolling bearings. *IEEE Transactions on Industrial Electronics*, *64*(8), 6506–6517. https://doi.org/10.1109/TIE.2017.2650873
Liu, S. Z., Sinha, R. S., & Hwang, S. H. (2021). Clustering-based noise elimination scheme for data pre-processing for deep learning classifier in fingerprint indoor positioning system. *Sensors*, *21*(13). https://doi.org/10.3390/s21134349
Liu, Y., Hu, Z., & Zhang, Y. (2021). Bearing feature extraction using multi-structure locally linear embedding. *Neurocomputing*, *428*, 280–290. https://doi.org/10.1016/j.neucom.2020.11.048
Mao, W., Ding, L., Tian, S., & Liang, X. (2020). Online detection for bearing incipient fault based on deep transfer learning. *Measurement: Journal of the International Measurement Confederation*, *152*, 107278. https://doi.org/10.1016/j.measurement.2019.107278
Mohanty, S., Gupta, K. K., & Raju, K. S. (2018). Hurst based vibro-acoustic feature extraction of bearing using EMD and VMD. *Measurement: Journal of the International Measurement Confederation*, *117*(February 2016), 200–220. https://doi.org/10.1016/j.measurement.2017.12.012
Nkikabahizi, C., Cheruiyot, W., & Kibe, A. (2022). Chaining Zscore and feature scaling methods to improve neural networks for classification. *Applied Soft Computing*, *123*, 108908. https://doi.org/10.1016/j.asoc.2022.108908
Peng, Y., & Dong, M. (2011). A hybrid approach of HMM and grey model for age-dependent health prediction of engineering assets. *Expert Systems with Applications*, *38*(10), 12946–12953. https://doi.org/10.1016/j.eswa.2011.04.091
Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, *77*(2), 257–286. https://doi.org/10.1109/5.18626
Seo, Y., Choi, Y., & Choi, J. (2017). River stage modeling by combining maximal overlap discrete wavelet transform, support vector machines and genetic algorithm. *Water (Switzerland)*, *9*(7), 525. https://doi.org/10.3390/w9070525
Tobon-Mejia, D. A., Medjaher, K., Zerhouni, N., & Tripot, G. (2010). A Mixture of Gaussians Hidden Markov Model for failure diagnostic and prognostic. *2010 IEEE International Conference on Automation Science and Engineering, CASE 2010*, *May 2014*, 338–343. https://doi.org/10.1109/COASE.2010.5584759
Viterbi, A. J. (1967). Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm. *IEEE Transactions on Information Theory*, *13*(2), 260–269. https://doi.org/10.1109/TIT.1967.1054010
Vrignat, P., Aggab, T., Avila, M., Duculty, F., & Kratz, F. (2019). Improvement indicators for Total Productive Maintenance policy. *Control Engineering Practice*, *82*(October 2018), 86–96. https://doi.org/10.1016/j.conengprac.2018.09.019
Vrignat, P., Kratz, F., & Avila, M. (2022). Sustainable manufacturing, maintenance policies, prognostics and health management: A literature review. *Reliability Engineering and System Safety*, *218*(October 2020), 108140. https://doi.org/10.1016/j.ress.2021.108140
Wang, H., Wang, D., Liu, H., & Tang, G. (2022). A predictive sliding local outlier correction method with adaptive state change rate determining for bearing remaining useful life estimation. *Reliability Engineering and System Safety*, *225*(January), 108601. https://doi.org/10.1016/j.ress.2022.108601
Wang, L., Cao, H., & Fu, Y. (2022). A bearing prognosis framework based on deep wavelet extreme learning machine and particle filtering. *Applied Soft Computing*, *131*, 109763. https://doi.org/10.1016/j.asoc.2022.109763
Wardhana, I., Gouiaa-Mtibaa, A., Vrignat, P., & Kratz, F. (2023). Data Preprocessing in the Context of PHM - Challenges and Solutions. *Proceedings of International Conference on Computers and Industrial Engineering, CIE*, *3*, 1523–1532.
Xie, S., Wang, X. X., Qu, C., Wang, X. X., & Guo, J. (2013). Impacts of different wind speed simulation methods on conditional reliability indices. *International Transactions on Electrical Energy Systems*, *20*(July 2013), 1–6. https://doi.org/10.1002/etep
Zhu, J., Chen, N., & Shen, C. (2020). A new data-driven transferable remaining useful life prediction approach for bearing under different working conditions. *Mechanical Systems and Signal Processing*, *139*, 106602. https://doi.org/10.1016/j.ymssp.2019.106602
Zhu, R., Peng, W., Wang, D., & Huang, C. G. (2022). Bayesian transfer learning with active querying for intelligent cross-machine fault prognosis under limited data. *Mechanical Systems and Signal Processing*, *183*(July 2022), 109628. https://doi.org/10.1016/j.ymssp.2022.109628