Check for updates

Received: 19 May 2023 | Revised: 26 September 2023 | Accepted: 28 September 2023

DOI: 10.1002/ese3.1597

ORIGINAL ARTICLE

Data-driven models for predicting remaining useful life of high-speed shaft bearings in wind turbines using vibration signal analysis and sparrow search algorithm

Ravi Pandit | **Weixun Xie**

School of Aerospace, Transport and Manufacturing (SATM), Centre for Life-Cycle Engineering and Management, Cranfield University, Cranfield, UK

Correspondence
Ravi Pandit, School of Aerospace, Transport and Manufacturing (SATM), Centre for Life-Cycle Engineering and Management, Cranfield University, Cranfield, UK.
Email: ravi.pandit@cranfield.ac.uk

Abstract

Wind turbine bearings play a crucial role in ensuring the safe and efficient operation of wind turbines. Accurate estimation of the remaining useful life (RUL) of bearings can significantly reduce operating and maintenance costs. In this paper, we propose three advanced data-driven models to predict the RUL of high-speed shaft bearings in wind turbines. These models combine the sparrow search algorithm (SSA) with three different regression models, namely support vector machine, random forest (RF) regression and Gaussian process regression. The models are based on features extracted from the vibration signal analysis, and the features are selected based on their monotonicity to evaluate bearing degradation. To optimize the performance of the regression models, all model parameters are tuned using the SSA algorithm. The proposed models are validated using vibration data collected from a real 2 MW commercial wind turbine. Our results demonstrate that the proposed models are effective in predicting the RUL of wind turbine bearings, and the SSA algorithm improves the accuracy of the predictions.

KEYWORDS

Gaussian process regression, random forest regression, remaining useful life estimation, sparrow search algorithm, support vector machine

1 | INTRODUCTION

Wind power is a renewable and clean source of energy, and with the ongoing climate crisis, many countries are turning to wind power generation facilities. The wind power industry has experienced significant growth in recent years, with wind turbines getting larger in size and installed at higher heights to capture stronger winds. According to data provided by the European Wind Energy Association (WWEA),[1] the installed wind power capacity reached 650.8 GW by the end of 2019. However, most of these turbines are located in remote mountainous areas or offshore, which poses significant challenges in transporting and hoisting the large components. The traditional method of waiting for system failure before replacing failed parts results in long downtime and puts enormous pressure on maintenance staff and standby parts inventory, increasing operating costs and ultimately affecting the Levelized Cost of Energy (LCoE).[2] The cost of operation and maintenance (O&M) is a significant

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2023 The Authors. *Energy Science & Engineering* published by the Society of Chemical Industry and John Wiley & Sons Ltd.

portion of the total annual wind turbine cost, with O&M accounting for 20%–25% of the LCoE.[3,4] To address this issue, improving the reliability of key components and predicting failures by improving maintenance strategies are currently the main ways to reduce O&M costs and lower the LCoE of wind power.

The growing size of wind turbines, their remote location, and the unpredictability of wind patterns necessitate efficient maintenance strategies. Maintenance strategies for wind turbines typically include corrective and proactive maintenance, the latter being further classified into scheduled and predictive maintenance.[5] Proactive maintenance is essential to prevent failures, and determining when maintenance is required requires knowledge about the current state of system degradation and an estimate of the Remaining Useful Life (RUL) of the system. RUL represents the useful life of an asset during a specific period of operation, and predicting it is a critical concept in Prediction and Health Management (PHM) to ensure high system availability over their life cycles gain lots of attention in recent years in addressing wind turbine challenges.[6] The data used in wind turbine RUL estimation can be categorized as recorded run-to-failure data and condition monitoring (CM) data.

CM data is particularly crucial for machines that cannot be allowed to run to failure. Any data associated with the degradation process, such as vibration, temperature, loading, rotating speed, current and voltage data, can theoretically be used to predict RUL. For example, Cheng et al.,[7] predicted gearbox fault and RUL by extracting features from generator current signals, while Sivalingam et al.,[8] used temperature variation signals to predict wind turbine power converter RUL. Hu et al.,[9] shows the distribution of temperature characteristic parameters of critical bearings in a wind turbine, which could be utilized to predict RUL after processing rotor speed and bearing temperature data obtained through the SCADA system.

Currently, most literature focuses on predicting the remaining life of wind turbine gearbox bearings. Vibration analysis is the most commonly used technique for monitoring wind turbines and diagnosing bearing faults. Vibration signals carry essential dynamic information related to fault characteristics and can be analyzed using time domain, frequency domain and time-frequency methods. Saidi et al.,[10] proposed technique to predict the RUL of wind turbine high-speed shaft bearings by extracting time domain indices from the spectral kurtosis of vibration signals. Furthermore, Djeziri et al.,[11] developed a physical model of wind turbines and collected data during normal and faulty operation for clustering. RUL is calculated in three steps: (1) projecting data in the remaining space using principal component analysis of normal operating data, (2) projecting all data into this remaining space resulting in normal and faulty clusters, and (3) calculating RUL as the ratio of Euclidean positions between the clusters and degradation velocity. Rommel et al.,[12] estimated wind turbine remaining life by evaluating load profiles at blade roots, rotor hub centre, and tower head using physical models.

Due to the availability of data and the rise of AI techniques, data-driven models have also been applied in recent years to estimate the RUL of wind turbines such as Pattern recognition methods (e.g., exponential degradation, stochastic filtering-based, Markov deterioration and so on), machine learning methods [e.g., artificial neural networks (ANN), support vector machines (SVM), Convolutional Neural Network (CNN) and so on]. Pan et al.,[13] used Gaussian process (GP) and Bayesian inference methods to predict bearing failures 1 month in advance by abstracting the state of bearing temperature residuals. Teng et al.,[14] used an artificial neural network to train data-driven models and then fit a polynomial curve to reflect the long-term degradation process of bearings. Hu et al.[9] further established a performance degradation model using temperature characteristic parameters and Wiener process. Author of[7] used an ANFIS-based PF method to predict RUL using one-phase stator current of the generator. Nielsen et al.,[15] calibrated a Markov deterioration model using past inspection data and updated it using real-time data. Rezamand et al.,[16] proposed a real-time prediction approach for wind turbine bearings in varying operating conditions using vibration signals and an adaptive Bayesian algorithm.

Wind turbines are complex systems with many components and structures, making it difficult to accurately build a physics-based model for predicting failures. Wang et al.,[17] proposed a hybrid prognostic prediction method for RUL of rolling element bearings, using a correlation vector machine regression with different kernel parameters for sparse representation of the bearing regression data, and then employing the exponential decay model combined with the Fraser distance for adaptive estimation. Sun et al.,[18] proposed a method for predicting the remaining service life of hybrid cutting tools based on empirical mode decomposition and back-propagation neural networks. SVMs have been shown to be effective in solving problems that consider structural risk and have better generalisation capabilities than traditional machine learning methods such as ANN.[19] Detailed recent research relevant to wind turbine maintenance optimisation can be found in Yang and colleagues.[20–22]

20500505, 2023, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ese3.1597 by Test, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1.1 | Novelty and contribution to knowledge

Wind turbines are known to experience high-speed shaft bearing failures that result in significant downtime and maintenance costs, ultimately leading to increased Levelized Cost of Energy (LCoE). To mitigate these costs, it is imperative to improve the reliability of key components or predict failures through better maintenance strategies. This study proposes a novel approach for predicting the RUL of high-speed shaft bearings in wind turbines using a regression model-based approach. We utilize a sparrow search algorithm (SSA) to optimize the regression model's parameters, thus improving the model's prediction accuracy. A series of time and frequency domain parameters are generated from the analysis of the vibration signal of a high-speed wind turbine shaft, and the regression model is trained after feature processing. The trained model is then used to predict the RUL of the high-speed shaft bearing. The effectiveness of our proposed method is confirmed through the analysis of vibration signals obtained from a real high-speed shaft bearing in a 2 MW wind turbine. To the best of our knowledge, this is the first study to apply the SSA algorithm to predict the RUL of wind turbine bearings. Our proposed models can help wind farm operators to optimize maintenance schedules and reduce downtime, ultimately leading to increased profitability and sustainability of wind energy production.

2 | DATA DESCRIPTION

In this section, detailed description of the data used for constructing RUL model for wind turbine high-speed shaft bearing discussed. The technical framework for the model is presented in Figure 1, which includes data collection, processing, feature selection, machine learning model training and RUL estimation. The data set used in this study was collected from a 2 MW wind turbine shaft in the United States.[23] Specifically, 50 vibration signals were recorded from 7 March to 25 April, each of which is 6 s in length. These signals were collected by a monitoring system after an inner race fault caused the failure of a high-speed shaft bearing that was driven by a 20-tooth pinion gear. Figure 2 shows the 50 raw vibration trends that were used in the analysis.

2.1 | Preprocessing

The vibration signal carries the basic dynamic information related to the fault characteristics. By using time and frequency domain analysis techniques we can obtain the underlying characteristics of the signal. In addition to the traditional statistical features in the time domain such as root mean square, kurtosis and so forth. Other statistical features in the frequency domain are also widely used for the prediction of the RUL of bearings. Spectral kurtosis (SK), for example, is considered to be a powerful tool that has proven to find successful applications in vibration-based CM.[24] It can indicate the presence of series of transients and their locations in the frequency domain. For the signal x, its SK value $K(f)$ can be calculated based on the short-time Fourier transform of the signal using the following equations.[25]

$$K(f) = \frac{\langle |S(t,f)|^4 \rangle}{\langle |S(t,f)|^2 \rangle^2} - 2, f \neq 0, \tag{1}$$

$$S(t,f) = \int_{-\infty}^{+\infty} x(t) w(t-\tau) e^{-2\pi f t} t, \tag{2}$$

where $w(t)$ is the window function used in STFT. Then SK is taken as signal x, and its mean standard deviation, std, Skewness and kurtosis are calculated.

2.2 | Feature selection

The process of feature selection was meticulously executed on the ensemble of 15 features extracted in

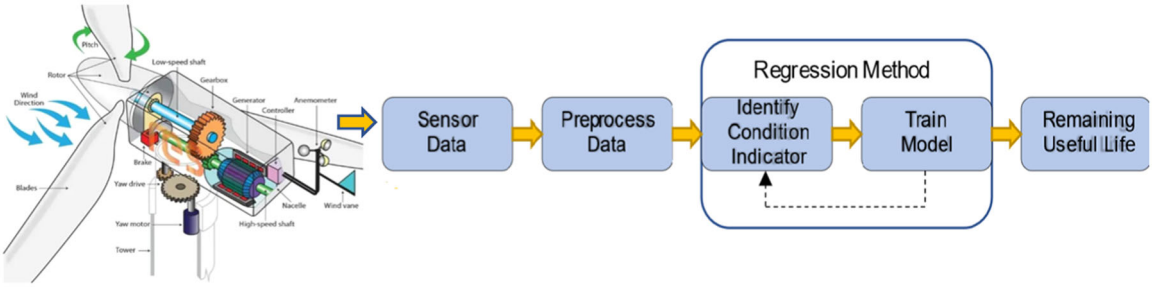


FIGURE 1 Proposed framework for wind turbine high-speed shaft bearing RUL prediction. RUL, Remaining Useful Life.

20500505, 2023, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ese3.1597 by Test, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

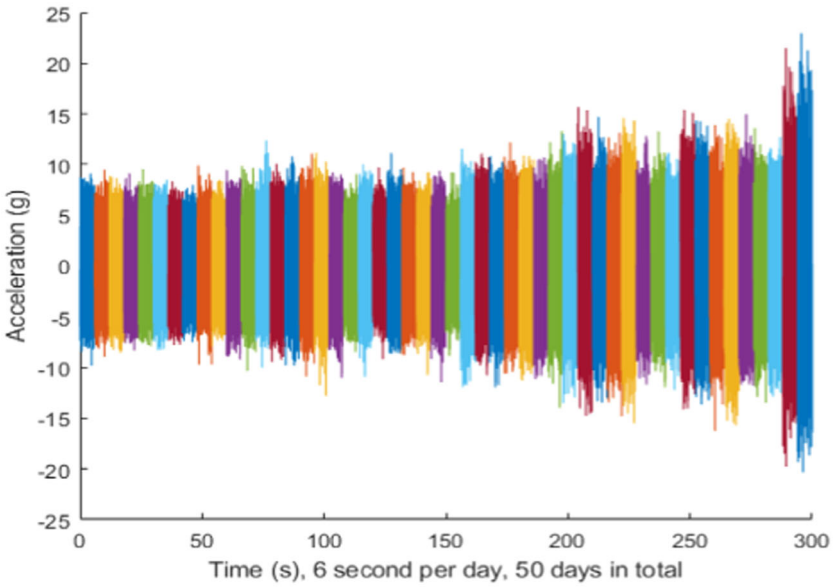


FIGURE 2 The vibration signals.

TABLE 1 Relevant time and frequency domain features.[26]

Features	Expression
Skewness	$\frac{1}{m}\sum_{i=1}^{m}\frac{(x_i-\bar{x})^3}{p^3}$
Kurtosis	$\frac{1}{m}\sum_{i=1}^{m}\frac{(x_i-\bar{x})^4}{p^4}$
Crest factor	$\frac{x_{\max}}{RMS}$
Shape factor	$\frac{RMS}{\frac{1}{m}\sum_{i=1}^{m}\lvert x_i\rvert}$
Impulse factor	$\frac{x_{\max}}{\frac{1}{m}\sum_{i=1}^{m}\lvert x_i\rvert}$
Margin factor	$\frac{x_{\max}}{\left(\frac{1}{m}\sum_{i=1}^{m}\lvert x_i\rvert\right)^2}$
SK mean	$\frac{1}{m}\sum_{i=1}^{m}skx_i$
SK std	$\left(\frac{1}{m}\sum_{i=1}^{m}skx_i-skmean^{sk}\right)$
SK skewness	$\frac{1}{m}\sum_{i=1}^{m}\frac{(skx_i-s\bar{k}x)^3}{p^3}$
SK kurtosis	$\frac{1}{m}\sum_{i=1}^{m}\frac{(skx_i-s\bar{k}x)^4}{p^4}$

the preceding section. The time and frequency domain features, pivotal to our analysis, are succinctly enumerated in Table 1, accompanied by their corresponding mathematical expressions. The representation of these features and their respective expressions is provided below.

To improve the prediction accuracy, it was necessary to filter out the features that did not contribute significantly to the RUL prediction model. The filtering process involved selecting the most important features based on their relevance to the RUL prediction. This was achieved by using a feature selection algorithm, which evaluated each feature based on its contribution to the prediction model. The algorithm ranked the features according to their relevance and selected the most important features for further analysis. By eliminating the less important features, the prediction accuracy of the model was significantly improved as shown in Figure 3. This illustrative visualization unveils the comprehensive evaluation of these features, revealing certain elements—such as SK kurtosis and SK skewness—that demonstrated less-than-anticipated efficacy in their predictive capacity. Conversely, the efficacy of kurtosis as a representative marker of the overarching degradation process emerged as particularly noteworthy.

To improve the accuracy of RUL estimation, it is important to carefully select and identify suitable indicators. Indicators that are suitable for RUL estimation should have certain characteristics, such as monotonicity, prognosability and trendability.

Monotonicity, as a defining characteristic, signifies an indicator's behaviour of consistently and uniformly increasing or decreasing in value as the machinery undergoes degradation. A well-defined monotonic indicator is instrumental in manifesting a distinct and coherent degradation trend, thereby substantially facilitating the accuracy of RUL predictions. This characteristic is formally defined by the following equation:

$$Monotonicity(x_i) = \left|\frac{positive(diff(x_i)) - negetive(diff(x_i))}{n-1}\right|, \quad (3)$$

where n is the number of measurement points, in this case $n = 50$.

20500505, 2023, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ese3.1597 by Test, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

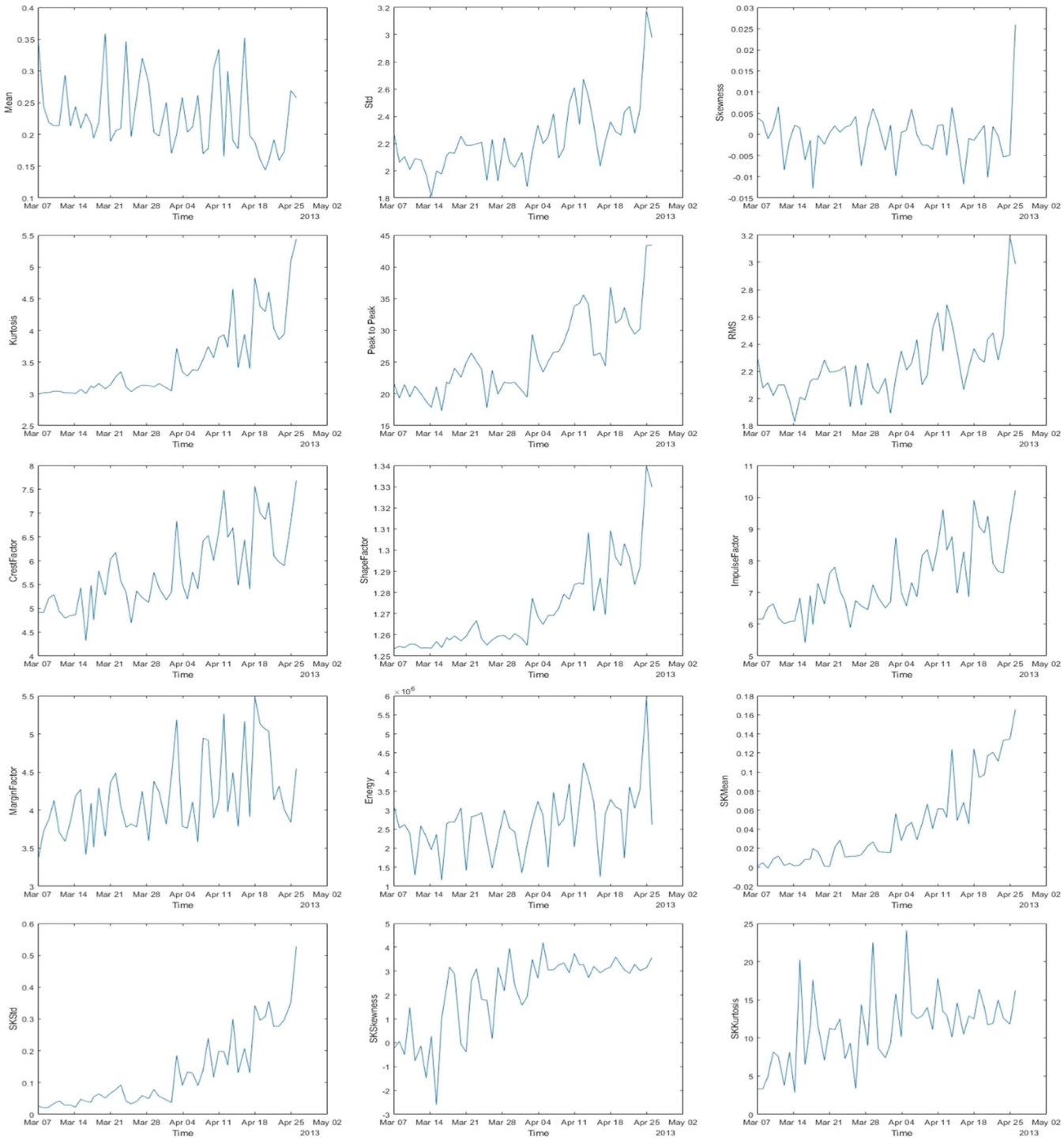


FIGURE 3 Evolution of 15 features over the 50 days.

The degradation of the bearings is an irreversible process, so for this study, monotonicity as the criterion of feature selection is chosen. Figure 4 shows the monotonicity metric of 15 features obtained in the previous section. To obtain the best prediction results, features with a score larger than 0.2 are selected for the next step.

3 | METHODS THEORETICAL DESCRIPTIONS

This paper employs three distinct regression models to predict the RUL of high-speed shaft bearings in wind turbines. The SVM regression, random forest regression (RFR) and Gaussian process regression (GPR)

20500505, 2023, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ese3.1597 by Test, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

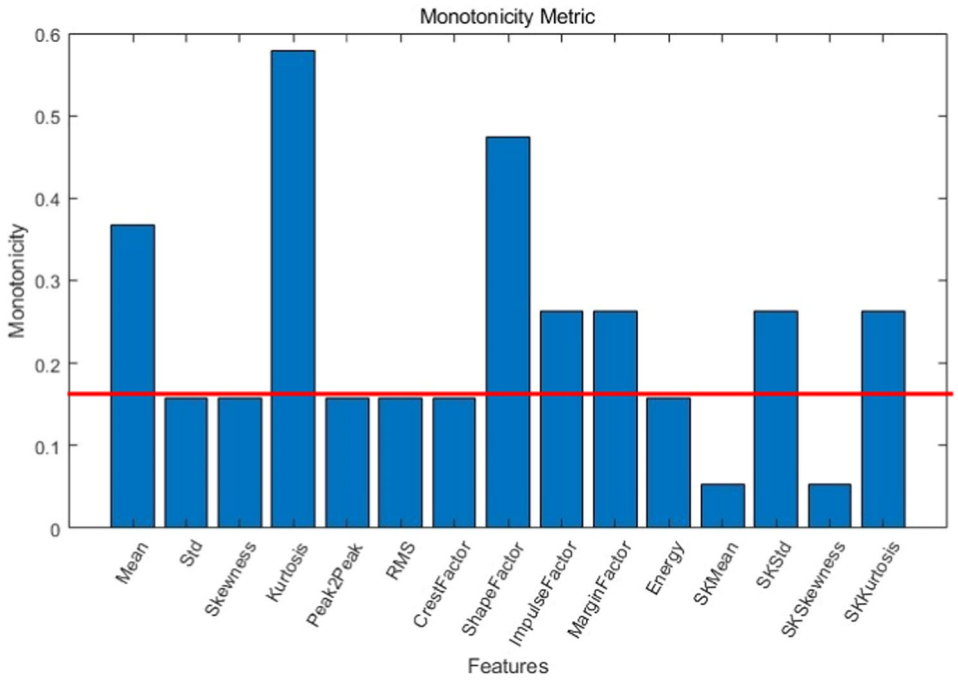


FIGURE 4 Monotonicity metric of features.

models are utilized. Furthermore, a SSA is implemented to fine-tune the model parameters during the training process. The following section provides a detailed description of these regression models and the optimization algorithm.

3.1 | SVM

SVM is a type of generalized linear classifier that utilizes supervised learning for binary data classification.[27] The decision boundary is set as the maximum-margin hyperplane over the learned samples. SVMs have gained popularity in engineering due to their effectiveness in handling nonlinear mapping problems and their ability to determine the final decision function by several support vectors. In situations where only a limited number of samples are available, SVMs are particularly useful in achieving this objective. As a result, SVMs are increasingly being employed for predicting RUL, owing to their efficacy in processing small data samples. The principle of SVM is: Given a sample set $S = \{(x_i, y_i)\ |_{i=1}^{n}, x_i \in R^n, y_i \in R\}$, where x_i is the input variable and y_i is the predicted value for, the regression function $f(x)$ is

$$f(x) = <w \bullet x> + b, \tag{4}$$

where w is the weight vector, $w \in R^n$; b is the bias threshold, $b \in R$. w and b can be obtained by solving the following optimal problem.

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*), \tag{5}$$

$$S.\ t. \begin{cases} <w \bullet x_i> + b - y_i \ll \xi_i + \varepsilon \\ y_i - <w \bullet x_i> - b \ll \xi_i^* + \varepsilon, \\ \xi_i, \xi_i^* \gg 0 \end{cases} \tag{6}$$

where C is the penalty factor, ξ_i and ξ_i^* are the relaxation factors and ε is the insensitivity factor.

When the data exhibit a nonlinear relationship with each other, a kernel function is introduced into the SVM to map the original data into a high-dimensional space, transforming the nonlinear problem into a linear problem to solve.[28] The RBF kernel function is the most commonly used, and its expression is:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{p^2}\right), \tag{7}$$

where p is kernel function indices.

The Lagrangian function is introduced to transform the optimization problem into a convex quadratic programming problem[27,28]:

20500505, 2023, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ese3.1597 by Test, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\max W\left(a_i, a_i^*\right) = -\frac{1}{2}\sum_{i,j}^{n}\left(a_i - a_i^*\right)\left(a_j - a_j^*\right) \bullet K\left(x_i, x_j\right) - \varepsilon\sum_{i=1}^{n}\left(a_i - a_i^*\right) + \sum_{i=1}^{n} y_i\left(a_i - a_i^*\right), \quad (8)$$

$$S.\ t.\ \begin{cases} \sum_{i=1}^{n}\left(a_i - a_i^*\right) = 0 \\ 0 < a_i < C \\ 0 < a_i^* < C \end{cases}, \quad (9)$$

where a_i, a_i^* is the Lagrangian operator and $W\left(a_i, a_i^*\right)$ is the function that solves for a_i, a_i^*. According to the KKT condition, the SVM regression function is:

$$f(x) = \sum_{i=1}^{n}\left(a_i - a_i^*\right)K\left(x_i, x\right) + b. \quad (10)$$

3.2 | Random forest

The RFR model is comprised of several regression trees that have no correlation with one another. The output of the model is ultimately decided by the collective decisions of all the decision trees in the forest. The randomness of the random forest is manifested in two distinct ways[29]:

- The randomness of the samples, where a specific number of samples are randomly selected from the training set to serve as the root node samples of each regression tree.
- The randomness of the features, where a specific number of candidate features are randomly chosen to construct each regression tree. The most appropriate features are then selected as split nodes.

The principle of the algorithm is described as follows.[30] Randomly draw m sample points from the training sample set $S$to obtain a new subtraining set $S_1, \ldots, S_n$. Then after using the subtraining set, train a CART regression tree (decision tree), here in the training process, the cut rule for each node is to select k features at random from all features, and then select the optimal cut point from these k features before dividing the left and right subtrees. Note that the obtained decision trees here are all binary trees. By the second step, many CART regression tree models can be generated and final prediction of each CART regression tree is the mean value of the leaf nodes reached by that sample point.

Key Attributes of Random Forests in Constructing Regression Trees: Sampling and Comprehensive Splitting Approach. The underpinning methodology of random forests entails two distinctive random sampling procedures. Specifically, the random forest algorithm conducts sampling across both the rows (samples) and columns (features) of the input data. In terms of sample sampling, a 'put-back' strategy is employed, thereby allowing for the potential inclusion of duplicates within the subset of samples generated through the sampling process.

Assuming that the input samples are N, then the samples sampled are also N. This makes it relatively easy to avoid over-fitting by not having all the samples in each tree during training, and then sampling the features, selecting $m\,(m << M)$ from the M features. This is followed by a regression tree using a full split on the sampled data. Each regression tree is a specialist in a narrow field (because we select m features from M features for each regression tree to learn), so that there are many experts in the random forest who are proficient in different fields, and a new problem (new input data) can be viewed from different perspectives, and eventually each expert will come up with his or her own result, which will be averaged. The base learner of a random forest is not a weak learner but a strong learner, consisting of a strong decision tree with a very high depth.

CART regression trees are based on the principle of minimum mean squared error (MSE). That is, for any division of feature A, corresponding to any division point s on both sides of the division into data sets D_1 andD_2, find the feature and eigenvalue division point that minimises the MSE of the respective sets of D_1 and D_2, while minimising the sum of the MSE of D_1 and D_2. The expressions are,

$$\min_{A,s}\left[\min_{c_1}\sum_{x_i \in D_1(A,s)}(y_i - c_1)^2 + \min_{c_2}\sum_{x_i \in D_2(A,s)}(y_i - c_2)^2\right], \quad (11)$$

where c_1 is the mean of the sample output for the D_1 data set and c_2 is the mean of the sample output for the D_2 data set. The prediction of the CART regression tree is based on the mean of the leaf nodes, so the prediction of the random forest is the average of the predicted values of all the trees.[29,30]

3.3 | GPR

GPR is a powerful nonparametric model utilized for regression analysis through a GP prior. The GPR model

20500505, 2023, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ese3.1597 by Test, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

assumptions consist of noise, or regression residuals, and a GP prior, which are resolved through Bayesian inference.[31] The process of solving GPR, also referred to as hyper-parameter learning, involves determining the hyper-parameters in the kernel function by learning samples in accordance with the Bayesian approach. By applying Bayes' theorem, the hyper-parameter posterior of GPR is expressed as a function of the likelihood, which is obtained by marginalizing the output of the GPR, and the prior distribution of the hyper parameters. From Bayes' theorem, the hyper-parameter posterior of GPR is expressed as follows[31]:

$$p(\theta|X,y) = \frac{p(y|X,\theta)p(\theta)}{p(y|X)}, \tag{12}$$

where θ denotes the hyperparameters of the GPR, including the hyperparameters of the kernel function and the variance of the residuals σ_n^2. $p(y|X,\theta)$ is the likelihood, which, as a non-parametric model, is the marginal likelihood obtained by marginalising the output of the GPR:

$$p(y|X,\theta) = \int p(y|f,X,\theta)p(f|X,\theta)df. \tag{13}$$

The GPR likelihood follows a normal distribution when the residuals adhere to an iid normal distribution using following equation.

$$p(y|X,\theta) = N\left(y \,\middle|\, 0, K + \sigma_n^2 I\right). \tag{14}$$

The GPR's optimal solution is a maximum likelihood estimate, which involves nonlinear optimization. when its common solution is a great likelihood estimate that includes nonlinear optimization.[4]

3.4 | SSA

A revolutionary swarm intelligence optimisation technique called the SSA was proposed in 2020 and was primarily motivated by the foraging and anti-predatory behaviour of sparrows.[32] In the study, SSA was employed to discover the most suitable hyperparameters for the proposed machine learning models, namely SVM, GPs, and RF. The primary goal of using the SSA was to fine-tune the hyperparameters of these models to align the characteristics of the data with the structure of each model. This process aimed to create prediction models for RUL that exhibit high levels of accuracy.

In the foraging behaviour of sparrows, there are two distinct roles, the finder and the follower. Finders usually have high energy reserves and are responsible for finding food-rich areas, and when they find food, they provide foraging directions for the entire species. Followers forage based on the information provided by the finder. In SSA, finders with better fitness values prioritise access to food during the search. During each iteration, the discoverer's position is updated as described below[33]:

$$X_{x,j}^{t+1} = \begin{cases} X_{x,j}^{t} \bullet \exp\left(\dfrac{-i}{\alpha \bullet iter_{\max}}\right) & if\ R_2 < ST \\ X_{x,j}^{t} + Q \bullet L & if\ R_2 > ST \end{cases}. \tag{15}$$

Where t represents the number of current iterations and $iter_{\max}$ is a constant indicating the maximum number of iterations. $X_{x,j}$ represents the position information of the i-th sparrow in the j-th dimension. $\alpha \in [0,1)$ is a random number. $R_2 \in [0,1]$ and $ST \in [0.5,1]$ denote the warning and safety value. Q is a random number that obeys a normal distribution. L denotes a $1 \times d$ matrix, where each element in the matrix is all 1. When $R_2 < ST$, this means that there are no predators around the foraging environment at this time and the finder can perform an extensive search operation. When $R_2 \geq ST$, this means that some sparrows in the population have spotted a predator and alerted the rest of the population, and all sparrows need to fly quickly to other safe places to forage. The updated description of the follower's location is as follows.

$$X_{x,j}^{t+1} = \begin{cases} Q \bullet \exp\left(\dfrac{X_{worst}^{t} - X_{x,j}^{t}}{i^2}\right) & if\ i > \dfrac{n}{2} \\ X_P^{t+1} + \left|X_{i,j}^{t} - X_P^{t+1}\right| \bullet A^{+} \bullet L & otherwise \end{cases}, \tag{16}$$

where X_P is the optimal position currently occupied by the discoverer and X_{worst} denotes the current global worst position. A^+ denotes a $1 \times d$ matrix where each element is randomly assigned a value of 1 or -1 and $A^+ = A^T(AA^T)^{-1}$. When $i > \frac{n}{2}$, this indicates that the i-th joiner with a low fitness value is not getting food and is in a very hungry state, at which point it needs to fly to elsewhere to forage for more energy.[32] When aware of challenges, sparrow populations engage in antipredatory behaviour, which is mathematically expressed as follows.

$$X_{x,j}^{t+1} = \begin{cases} X_{best}^{t} + \beta \bullet \left|X_{x,j}^{t} - X_{best}^{t}\right| & if\ f_i > f_g \\ X_{x,j}^{t} + K \bullet \left(\dfrac{X_{x,j}^{t} - X_{worst}^{t}}{(f_i - f_w) + \varepsilon}\right) & if\ f_i = f_g \end{cases}, \tag{17}$$

20500505, 2023, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ese3.1597 by Test, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where X_{best} is the current global optimum. β, the step control parameter, is a random number with a normal distribution with mean 0 and variance 1. $K \in [-1,1]$ is a random number and f is the current fitness value of the individual sparrow. f_g and f_w are the current global best and worst fitness values, respectively. ε is the smallest constant to avoid zero in the denominator.

For simplicity, when $f_i > f_g$ means that the sparrow is at the edge of the population and is extremely vulnerable to predators. $f_i = f_g$ means that sparrows in the middle of the population are aware of the danger and need to move closer to other sparrows to minimise their risk of predation. K indicates the direction of movement of the sparrow and is also the step control parameter.

The concept behind hyperparameter tuning is that adjusting certain settings or parameters of a machine learning model can significantly impact its performance and ability to make accurate predictions. The SSA, in this case, served as the mechanism for systematically exploring and determining the optimal combination of hyperparameters for the SVM, GP and RF models.

To achieve improved training outcomes, it is imperative to divide the training set in the most random manner possible while ensuring that the data incorporates all phases of the degradation process. Furthermore, the model parameters are optimized using a SSA, and the optimal parameter settings of the SSA are obtained after several debugging iterations.[30,31] In this study, 80% and 70% of the available data were selected as the training set to train the model, and the results will be elaborated upon in the subsequent section.

4 | RESULT AND DISCUSSIONS

Mean Absolute Error is used to evaluate the training process. MAE, which may accurately depict the actual situation of the predicted value error and be used as an evaluation metric for the training model's performance, is the average of the absolute error between the true value and the predicted value. Its expression is:

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|p(y_i) - y_i|, \tag{18}$$

where m is the number of training samples, y_i is the expected RUL of the i-th sample and $p(y_i)$ is the predicted RUL of the i-th sample. The final training results are shown in Table 2.

Here, MAE is not a criterion for evaluating the merits of the three models. The value of MAE is only used as a criterion for evaluating the same model after each training, the smaller the value of MAE, the better the training result of the model. Since the GRP model is different from the other two, the training results are not evaluated here. After selecting the best training results, the three trained regression models were tested using the same test set to compare the predicted remaining lifespan with the actual lifespan. The results are shown below. Figure 5 shows the prediction results of the model trained with 70% of the data as the training set, and the following three figures show the prediction results of the model trained with 80% of the data as the training set.

TABLE 2 Models training results.

Models	70%	80%
SVM	1.46	1.87
RF	2.31	2.36
GPR	\	\

Abbreviations: GPR, Gaussian process regression; RF, random forest; SVM, support vector machines.

Figure 5 demonstrates that although each model deviates somewhat from the genuine RUL curve, they can all forecast the remaining life to some extent. We offer four parameters to evaluate the regression models: MSE, RMSE, MAE, and R^2. To more accurately assess each regression model's performance. The other three metrics are shown below, and the definition of MAE has already been discussed in the evaluation of the results of the model training.

MSE is a popular error metric for regression models, which measure the variance of the residuals. It represents the mean of the squared differences between the original and predicted values in the data set and is calculated as:

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(p(y_i) - y_i)^2, \tag{19}$$

where m is the number of training samples, y_i is the expected RUL of the i-th sample and $p(y_i)$ is the predicted RUL of the i-th sample.

Root Mean Squared Error (RMSE) is an extension of the MSE, which is the square root of MSE. It measures the standard deviation of residuals and is calculated as:

$$RMSE = \sqrt{MSE}. \tag{20}$$

The R-squared, also known as the coefficient of determination, is the square of the correlation coefficient R. It is an important statistical parameter in a linear regression model. It indicates how well the regression

20500505, 2023, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ese3.1597 by Test, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

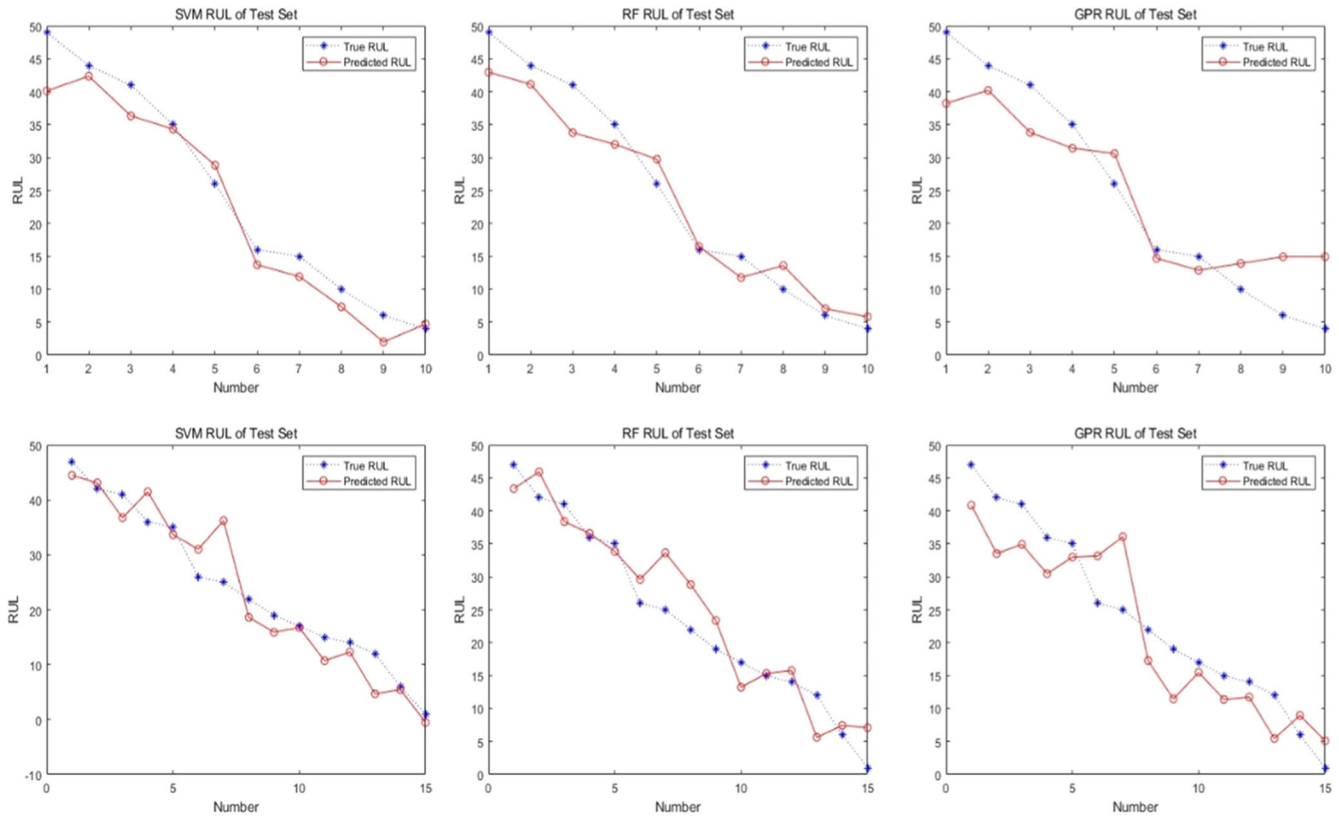


FIGURE 5 RUL estimation results. RUL, Remaining Useful Life.

model fits the actual values, with R close to 1 indicating a better fit, and is expressed as:

$$R^2 = 1 - \frac{\sum(y_i - p(y_i))^2}{\sum(y_i - \bar{y})^2}, \tag{21}$$

where y_i is the expected RUL of the i-th sample, $p(y_i)$ is the predicted RUL of the i-th sample and $\bar{y}$ is the mean value of y.

The results indicated that the SVM model achieved the highest accuracy among the three, suggesting that the fine-tuned hyperparameters obtained through the SSA allowed the SVM model to more effectively capture the underlying patterns and relationships in the data. Following the SVM model, RF model demonstrated the second-best accuracy, while the GPs model achieved comparatively lower accuracy. Tables 3 and 4 show the evaluation parameters for the models obtained by training with 80% of the data as the training set and the model obtained by training with 70% of the data as the training set, respectively. It can be seen that the models obtained using 80% of the data as the training set are more accurate than those obtained using 70% of the data. When 80% of the data was used as the training set, the coefficients of determination of the models obtained were all greater than 0.9, at which point the SVM model reached a coefficient of determination of 0.97, which can indicate that the SVM can predict the remaining life of this

TABLE 3 Evaluation parameters for models when training data sets is 80%.

Models	MSE	RMSE	MAE	R^2
SVM	15.00	3.87	3.14	0.97
RF	14.76	3.84	3.30	0.93
GPR	27.13	5.21	0.20	0.90

Abbreviations: GPR, Gaussian process regression; RF, random forest; SVM, support vector machines.

TABLE 4 Evaluation parameters for models when training data sets is 70%.

Models	MSE	RMSE	MAE	R^2
SVM	20.51	4.53	3.54	0.91
RF	19.21	4.38	3.68	0.90
GPR	34.94	5.91	0.51	0.83

Abbreviations: GPR, Gaussian process regression; RF, random forest; SVM, support vector machines.

20500505, 2023, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ese3.1597 by Test, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

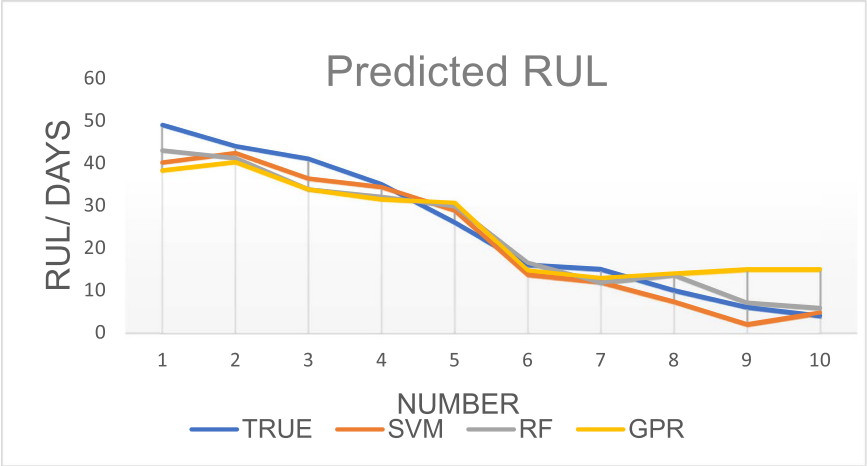


FIGURE 6 RUL estimation models performance comparison when 80% of the data is the training set. GPR, Gaussian process regression; RF, random forest; RUL, Remaining Useful Life; SVM, support vector machines.

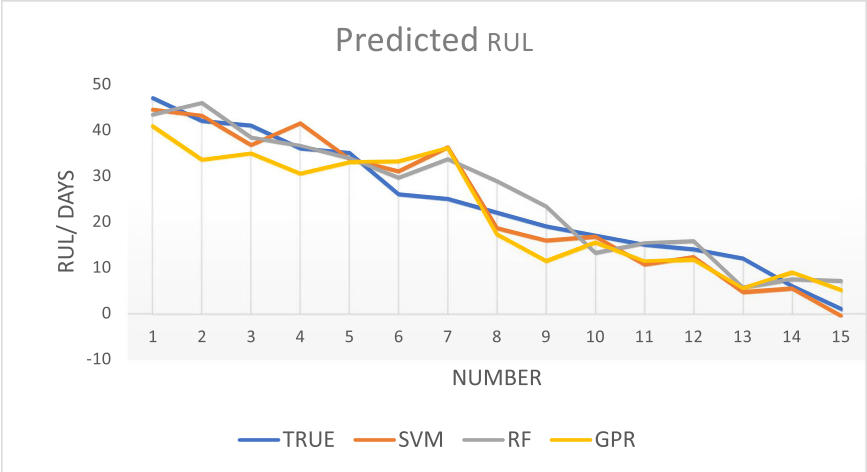


FIGURE 7 RUL estimation models performance comparison when 70% of the data is the training set. GPR, Gaussian process regression; RF, random forest; RUL, Remaining Useful Life; SVM, support vector machines.

bearing very well. In both cases, the SVM regression model is rated better than the other two regression models, followed by RF. when the regression curves of the three models are placed in the same graph, it can also be seen that the SVM regression curve is closer to the true value.

SVM strengths in handling complex data distributions, regularization, high-dimensional data, hyperparameter sensitivity, and balancing bias-variance trade-offs likely contributed to its superior performance in our study. Furthermore, the synergy between the SSA's optimization capabilities and SVM's sensitivity to hyperparameters, complex interactions, and data distribution adaptation likely resulted in SVM's significant outperformance over RF and GPs in our study. While all three algorithms benefited from SSA's optimization, SVM's unique characteristics allowed it to derive maximum advantage from the fine-tuned hyperparameters produced by the algorithm as shown in Figures 6, 7 and Tables 3 and 4.

5 | CONCLUSION

In conclusion, predicting the RUL of wind turbine high-speed shaft bearings has always been a daunting task. This paper addresses this challenge by proposing three highly promising methods for RUL prediction based on vibration signal analysis with feature selection based on monotonicity to evaluate the degradation of high-speed shaft bearings. Our results demonstrate that the proposed feature extraction method is highly effective. Furthermore, we leverage a SSA to optimize the parameters of a SVM, RFR model, and MAE values to evaluate the model training outcomes. These models are validated using vibration data collected from a real 2 MW commercial wind turbine. Our findings reveal that the proposed models combined with SSA algorithm are efficient in predicting the RUL, and the performance of the models improves with an increase in the proportion of the training set to the data set. Among the three models proposed, SVM model, after optimization of parameters

20500505, 2023, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ese3.1597 by Test, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

using the SSA algorithm, outperforms the others. Proposed research of this paper demonstrates the potential of these methods for RUL prediction of wind turbine high-speed shaft bearings and can have significant implications for the wind power industry.

ORCID

Ravi Pandit http://orcid.org/0000-0001-6850-7922

REFERENCES

1. European Wind Energy Association (EWEA). Wind report; 2016. Accessed May 1, 2016. https://www.ewea.org/
2. Pandit R, Infield D, Santos M. Accounting for environmental conditions in data-driven wind turbine power models. *IEEE Trans Sustain Energy* 2023;14(1):168-177. doi:10.1109/TSTE.2022.3204453
3. Operation and maintenance costs of wind-generated power, wind energy the fact. Accessed May 1, 2023. https://www.wind-energy-the-facts.org/
4. Pandit RK, Astolfi D, Durazo Cardenas I. A review of predictive techniques used to support decision making for maintenance operations of wind turbines. *Energies*. 2023;16(4):1654. doi:10.3390/en16041654
5. Cevasco D, Koukoura S, Kolios AJ. Reliability, availability, maintainability data review for the identification of trends in offshore wind energy applications. *Renew Sustain Energy Rev*. 2021;136:110414.
6. Hadjoudj Y, Pandit RK. Improving O&M decision tools for offshore wind farm vessel routing by incorporating weather uncertainty. *IET Renew Power Gener*. 2023;17:1488-1499. doi:10.1049/rpg2.12689
7. Cheng F, Qu L, Qiao W. Fault prognosis and remaining useful life prediction of wind turbine gearboxes using current signal analysis. *IEEE Trans Sustain Energy*. 2018;9(1):157-167. doi:10.1109/TSTE.2017.2719626
8. Sivalingam K, Sepulveda M, Spring M, Davies P. A review and methodology development for remaining useful life prediction of offshore fixed and floating wind turbine power converter with digital twin technology perspective. In: Proceedings 2018 2nd International Conference on Green Energy and Applications (ICGEA);2018:197-204. doi:10.1109/ICGEA.2018.8356292.
9. Hu Y, Li H, Shi P, et al. A prediction method for the real-time remaining useful life of wind turbine bearings based on the Wiener process. *Renew Energy*. 2018;127:452-460. doi:10.1016/j.renene.2018.04.033
10. Saidi L, ben Ali J, Bechhoefer E, Benbouzid M. Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived indices and SVR. *Appl Acoustics*. 2017;120:1-8. doi:10.1016/j.apacoust.2017.01.005
11. Djeziri MA, Benmoussa S, Sanchez R. Hybrid method for remaining useful life prediction in wind turbine systems. *Renew Energy*. 2018;116:173-187. doi:10.1016/j.renene.2017.05.020
12. Rommel DP, di Maio D, Tinga T. Calculating wind turbine component loads for improved life prediction. *Renew Energy*. 2020;146:223-241. doi:10.1016/j.renene.2019.06.131
13. Pan Y, Hong R, Chen J, Singh J, Jia X. Performance degradation assessment of a wind turbine gearbox based on multi-sensor data fusion. *Mech Mach Theory*. 2019;137:509-526. doi:10.1016/j.mechmachtheory.2019.03.036
14. Teng W, Zhang X, Liu Y, Kusiak A, Ma Z. Prognosis of the remaining useful life of bearings in a wind turbine gearbox. *Energies*. 2017;10(1):32. doi:10.3390/en10010032
15. Nielsen J, Sørensen J. Bayesian estimation of remaining useful life for wind turbine blades. *Energies*. 2017;10(5):664. doi:10.3390/en10050664
16. Rezamand M, Kordestani M, Orchard ME, Carriveau R, Ting DSK, Saif M. Improved remaining useful life estimation of wind turbine drivetrain bearings under varying operating conditions. *IEEE Trans Ind Inform*. 2021;17(3):1742-1752. doi:10.1109/TII.2020.2993074
17. Wang B, Lei Y, Li N, Li N. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Trans Reliab*. 2020;69:401-412.
18. Sun H, Cao D, Zhao Z, Kang X. A hybrid approach to cutting tool remaining useful life prediction based on the Wiener Process. *IEEE Trans Reliab*. 2018;67:1294-1303.
19. Zhang L, Liu Z, Luo D, Li J, Huang H, Review of remaining useful life prediction using support vector machine for engineering assets. In: 2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE);2013:1793-1799. doi:10.1109/QR2MSE.2013.6625925
20. Yang L, Chen Y, Ma X. A state-age-dependent opportunistic intelligent maintenance framework for wind turbines under dynamic wind conditions. *IEEE Trans Ind Inform*. 2023;19(10):10434-10443. doi:10.1109/TII.2023.3240727
21. Yang L, Wei F, Qiu Q. Mission risk control via joint optimization of sampling and abort decisions. *Risk Anal*. 2023:1-20. doi:10.1111/risa.14187
22. Yang L, Chen Y, Ma X, Qiu Q, Peng R. A prognosis-centered intelligent maintenance optimization framework under uncertain failure threshold. *IEEE Trans Reliab*. 2023:1-16. doi:10.1109/TR.2023.3273082
23. Bechhoefer E, Van Hecke B, He D. Processing for improved spectral analysis. Annual Conference of the Prognostics and Health Management Society, New Orleans, LA; 2013.
24. Xu M, Baraldi P, Yang Z, Zio E. A two-stage estimation method based on conceptors-aided unsupervised clustering and convolutional neural network classification for the estimation of the degradation level of industrial equipment. *Expert Syst Appl*. 2023;213:118962. doi:10.1016/j.eswa.2022.118962
25. Aizpurua JI, McArthur SDJ, Stewart BG, Lambert B, Cross JG, Catterson VM. Adaptive power transformer lifetime predictions through machine learning and uncertainty modeling in nuclear power plants. *IEEE Trans Ind Electron*. 2019;66:4726-4737.
26. Altaf M, Akram T, Khan MA, Iqbal M, Ch MMI, Hsu C-H. A new statistical features based approach for bearing fault diagnosis using vibration signals. *Sensors*. 2022;22:2012. doi:10.3390/s22052012
27. Boser BE, Guyon IM, Vapnik V. A training algorithm for optimal margin classifiers. *In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory—COLT*, Pittsburgh, PA, USA;1992:144.

20500505, 2023, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ese3.1597 by Test, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

28. Pandit R, Kolios A. SCADA data-based support vector machine wind turbine power curve uncertainty estimation and its comparative studies. *Appl Sci*. 2020;10:8685. doi:10.3390/app10238685
29. Lahouar A, Ben Hadj Slama J. Hour-ahead wind power forecast based on random forests. *Renew Energy*. 2017;109:529-541. doi:10.1016/j.renene.2017.03.064
30. Freund Y. Boosting a weak learning algorithm by majority. *Inform Comput*. 1995;121(2):256-285. doi:10.1006/inco.1995.1136
31. Rasmussen C, Williams C. *Gaussian Processes for Machine Learning*. The MIT Press; 2006.
32. Xue J, Shen B. A novel swarm intelligence optimization approach: sparrow search algorithm. *Syst Sci Control Eng*. 2020;8(1):22-34.
33. Thirunavukkarasu M, Sawle Y, Lala H. A comprehensive review on optimization of hybrid renewable energy systems using various optimization techniques. *Renew Sustain Energy Rev*. 2023;176:113192.

How to cite this article: Pandit R, Xie W. Data-driven models for predicting remaining useful life of high-speed shaft bearings in wind turbines using vibration signal analysis and sparrow search algorithm. *Energy Sci Eng*. 2023;11:4557-4569. doi:10.1002/ese3.1597

20500505, 2023, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ese3.1597 by Test, Wiley Online Library on [29/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



https://dspace.org/

Data-driven models for predicting remaining useful life of high-speed shaft bearings in wind turbines using vibration signal analysis and sparrow search algorithm

Pandit, Ravi; Xie, Weixun

2023-12-01
Wiley
https://doi.org/10.1002/ese3.1597
Pandit R, Xie W. (2023) Data-driven models for predicting remaining useful life of high-speed shaft bearings in wind turbines using vibration signal analysis and sparrow search algorithm. Energy Science & Engineering, Volume 11, Issue 12, December 2023, pp. 4557-4569

Downloaded from a DSpace repository. https://dspace.org/