Received: 8 May 2024 | Revised: 20 August 2024 | Accepted: 27 August 2024

DOI: 10.1002/ese3.1911

ORIGINAL ARTICLE

Comparative analysis of novel data-driven techniques for remaining useful life estimation of wind turbine high-speed shaft bearings

Ravi Pandit[1] | **Matilde Santos**[2] | **Jesus Enrique Sierra-García**[3]

[1]Centre for Life-Cycle Engineering and Management, Cranfield University, Wharley End, UK

[2]Computer Science and Physics, Complutense University of Madrid, Madrid, Spain

[3]Electromechanical Engineering Department, University of Burgos, Burgos, Spain

Correspondence
Ravi Pandit, Centre for Life-Cycle Engineering and Management, Cranfield University, Wharley End, UK.
Email: ravi.pandit@cranfield.ac.uk

Funding information
This research is supported by the Department for Science, Innovation & Technology (DSIT), UK under Tactical Fund Programme

Abstract

As the global momentum for wind power generation accelerates, the industry faces substantial challenges due to premature failures in wind turbine components. These failures, particularly in critical elements like the high-speed shaft bearing, lead to significant operational losses, including unplanned downtime and elevated maintenance costs. To mitigate these issues, it's crucial to have precise predictions of the remaining useful life (RUL) of these components, enabling timely interventions and more efficient maintenance schedules. This article proposes advanced, data-driven approaches for estimating the RUL of wind turbine high-speed shaft bearings, utilizing cutting-edge techniques such as long short-term memory (LSTM), bidirectional LSTM (BiLSTM), gated recurrent units (GRU), and random forest (RF) algorithms. Our analysis leverages vibration data from a 2 MW wind turbine equipped with a 20-tooth pinion gear, providing a thorough validation and comparison of these methodologies against traditional models. Our results reveal that the LSTM and BiLSTM models excel in both accuracy and computational efficiency for predicting RUL and enhancing system prognosis, surpassing the performance of conventional RF and GRU methods. This research underscores the potential of our innovative data-driven strategies to develop effective RUL estimation algorithms, significantly advancing wind turbine proactive operation and maintenance operations.

KEYWORDS

data-driven, deep learning, feature extraction, machine learning, remaining useful life estimation, wind turbine

1 | INTRODUCTION

The global demand for renewable energy has been on the rise, with corporations and individuals increasingly adopting environmentally friendly alternatives to fossil fuels. Governments around the world are setting ambitious targets to transition to renewable energy as their primary energy source in the near future. In the UK alone, renewable sources contributed to a remarkable 43.1% of electricity generation in 2021,[1] underlining the growing significance of this sustainable energy sector. Among renewable sources, wind power has emerged as a dominant player, accounting for over half of the total renewable electricity generated in 2020.

While wind turbines are a marvel of engineering and provide reliable power generation, the eventual breakdown of critical components is inevitable. These breakdowns can lead to catastrophic consequences, including prolonged downtime and substantial profit losses.[2] Wind farms, operating round-the-clock and spread across multiple locations, demand meticulous maintenance to avoid costly unexpected failures. These wind turbines are usually one of many in a wind farm, which operates typically 24/7 and is distributed in multiple locations. Some of these locations may not even be on land, just as many offshore wind farms now operating. The reason for offshore wind farms is that offshore wind generation shows tremendous potential compared to its onshore counterpart due to larger blades and stronger winds, among other advantages.

The remaining useful life (RUL) of wind turbines refers to the remaining duration during which the turbine can operate at its optimal efficiency and generate the expected power output. It represents the time left before any significant decrease in performance or potential failure may occur, leading to reduced power generation and operational effectiveness. This figure is subject to one or more variables, making it a critical component of condition-based maintenance and health monitoring techniques that help to mitigate unexpected and catastrophic breakdowns. Such breakdowns can significantly decrease a unit's expected profit margins, and the length of downtimes may be extended depending on the severity and suddenness of the failure.[3] For example, according to [4] repairing a single wind blade can cost up to $30,000, while a replacement can cost up to $200,000. Because of its geographical distribution, size, and complexity, repairing an unexpected wind turbine breakdown can be prohibitively expensive and time-consuming. These difficulties are compounded in the case of offshore (floating) wind turbines, owing to logistics and higher load factors.[5] Thus, accurate RUL estimation plays a crucial role in condition-based maintenance and health monitoring techniques, enabling timely inspections and efficient spare parts management. Mitigating failures and minimizing downtime are vital to ensure economic viability and safety.

The remainder of this paper is structured as follows: Section II provides a review of related works. Section III outlines the data-driven methodology and software setup used in our study. Section IV discusses the data set and pre-processing steps. In Section V, we present the results and performance comparisons with existing techniques, highlighting the strengths and weaknesses of our proposed models. Finally, Section VI presents our conclusions and outlines future research directions.

2 | LITERATURE REVIEW AND RESEARCH PROBLEM

Offshore wind turbines experience higher loads than their onshore counterparts due to the higher wind speeds off the coasts and the waves that impact floating turbines. Therefore, it is critical to developing a means of predicting a wind turbine's RUL to determine when it is expected to fail. Despite the constant research and publication of papers on wind turbines, the technology behind wind energy generation and the state-of-the-art methods available for addressing operation and maintenance (O&M) issues continue to evolve as wind energy production expands. As a result, older papers may contain outdated information due to the lack of resources and technological advancements that we have today, and our primary source of citations would be those published after 2015. The RUL concept has been studied in various fields, not only in wind turbines. RUL estimation can be used for many moving machinery and components due to the similarities between them, making it a cross- field application.

Deep learning is becoming increasingly popular in predicting the RUL because of its ability to automatically create a mapping relationship between raw data and the associated RUL. Among the deep learning models, convolutional neural networks are gaining popularity due to their powerful capacity to handle time-series information, as demonstrated in recent studies. Researchers have presented various approaches to predict the RUL of gearboxes, including a rule-based approach, a damage curve and rule-based hybrid approach, an approach based on quadratic discriminant analysis, and a combination of these approaches. In addition to these approaches, machine learning algorithms such as artificial neural network (ANN), support vector machine, and logistic regression have also been tested in predicting gearbox failure. A statistical particle filter-based fault prognostic and RUL estimation method for a wind turbine gearbox has been proposed, which uses two data-driven models and a forward-backwards filter to filter out the high-frequency sampling noise in the current signal. The proposed method accurately predicted the RUL roughly 9 h 30 min (560 min) before the actual failure time (within 20% confidence intervals).

Ding et al.[6] introduced a data-driven prognostic scheme for bearings using a novel health indicator combined with a gated recurrent unit (GRU) network. Their approach demonstrated improved prediction accuracy for RUL by effectively capturing the degradation trends in bearings. Another study by Meddour et al.[7] employed gray relational analysis for feature selection in a vibration data set, which was then used in an Adaptive-network-based fuzzy inference system to estimate the RUL of wind turbine bearings. Their method

showed significant improvements in prediction accuracy by identifying the most relevant features for model training. Similarly, Teng et al.[8] proposed a dynamic model-assisted RUL prediction for bearings using a cross-domain transformer network. This model leveraged domain adaptation techniques to enhance prediction robustness across different operational conditions. Sanchez et al.[9] proposed a stiffness degradation model for optimizing the blades on a wind turbine to achieve a longer life cycle and estimate RUL. Their RUL model is influenced by rotor speed, applied torque, and wind speed at the turbine's hub height. In another research, Pan et al.[10] proposed an inverse Gaussian degradation model with a random effect for RUL estimation, tested on the degradation data of a laser device. Zhang et al.[11] developed a novel approach using long short-term memory (LSTM) networks to predict RUL for machines. Their method addressed data uncertainty and environmental disturbances by incorporating a bidirectional LSTM, which processes input sequences in both forward and backward directions, thus enhancing prediction reliability.

The estimation of the RUL in wind turbines is crucial for extending their lifespan, optimizing performance, and reducing operational costs. This significance is well-established in existing literature. In recent years, data-driven algorithms have been increasingly applied to robustly estimate RUL in various wind turbine components. Notable works, such as those by B. Wang[12] and Ma et al.[13] have proposed hybrid prognostic models and convolution-based long short-term memory (CLSTM) networks for predicting RUL in rotating machinery using vibration data. Similarly, Liu et al.[14] explored the use of the generalized Cauchy process to enhance gearbox RUL estimation accuracy. Despite these advancements, there remains a limited exploration of variant recurrent neural network (RNN) techniques, such as LSTM, bidirectional LSTM (BiLSTM), and GRU, specifically for wind turbine RUL estimation. Addressing this knowledge gap, our research introduces novel data-driven algorithms and provides a comprehensive comparison with existing methods for estimating the RUL of high- speed shaft bearings in wind turbines. Leveraging LSTM, BiLSTM, GRU, and machine learning (random forest [RF]), our approach demonstrates superior accuracy and computational efficiency. A key innovation of this study is the thorough evaluation and comparison of these data-driven techniques, tailored specifically for wind turbine components. Unlike previous studies, our research provides a detailed analysis of how these advanced models perform relative to each other, highlighting their strengths and weaknesses in practical applications.

Figure 1 illustrates the flow chart of the proposed methodology for estimating the RUL of wind turbine high-speed shaft bearings. The process begins with data description and filtration, followed by feature extraction and analysis. The next step involves designing and developing models using RNNs such as LSTM, BiLSTM, GRU, and RF. These models are then trained and evaluated using performance metrics like MASE, mean absolute percentage error (MAPE), and root mean squared error (RMSE). The final stage involves comparing the results, highlighting the strengths of the proposed models in predicting RUL and system prognosis.

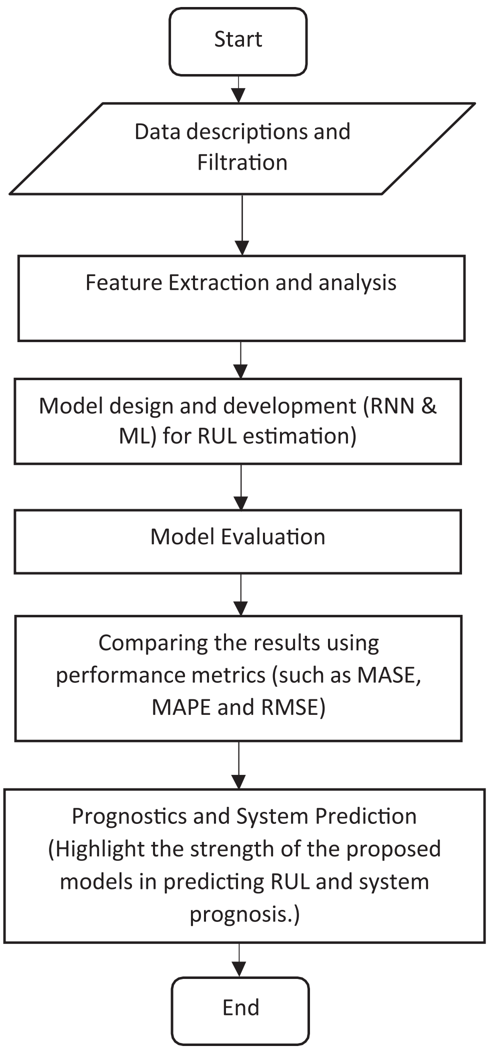


FIGURE 1 Flow chart of the proposed methodology.

3 | METHODOLOGIES AND EXPERIMENTAL SETUP

3.1 | LSTM theoretical descriptions

The LSTM is a type of recursive neural network inspired by the biological architecture of the brain. Unlike traditional feed-forward neural networks, LSTMs allow errors to be backpropagated over time across the layers of the network, resulting in a more consistent error rate and enabling the network to learn for longer periods of time.[15] As a result, LSTMs outperform both RNNs and traditional feed-forward neural networks and can automatically learn hierarchical patterns in deep structures, both in supervised and unsupervised learning scenarios. A more detailed explanation of LSTM theory can be found in [16].

In this research, we employed three different types of recurrent neural networks (RNNs) as well as an additional machine learning technique to estimate RUL. These techniques were LSTM, bidirectional LSTM, GRU, and RF. RNNs are particularly useful for analyzing time series and sequential data, as they can use their previous outputs as inputs for the next step, thereby exploiting the memory property of the network to improve future predictions. However, basic RNNs have a short-term memory problem due to the vanishing gradient problem. The vanishing gradient problem arises because, as we backpropagate through time, the weights in the earlier network layers can become extremely small, causing the calculated gradient to also become very small. This means that the update on the weights will have very little impact as it passes through the network, particularly at earlier layers, leading to a loss of information about longer sequences.

An LSTM (Figures 2 and 3)[17] incorporates internal mechanisms called gates, including the input gate (It), output gate (Ot), and forget gate (Ft), which allow for careful updating and control of the cell states (Ct). The input gate regulates new information that is passed into the cell state, while the forget gate determines which past information should be forgotten and not passed into the cell state. By selectively retaining and discarding information, the network can learn from important data and effectively handle long-term dependencies.

Unlike traditional RNNs, LSTM overcomes the vanishing gradient problem by employing additive properties for the cell state gradients, as well as the forget gate, which enables the network parameters to be updated without vanishing gradients. Meanwhile, a GRU also introduces the concept of gates to the network, but with fewer parameters than an LSTM due to the absence of the output gate, leading to more efficient computation.

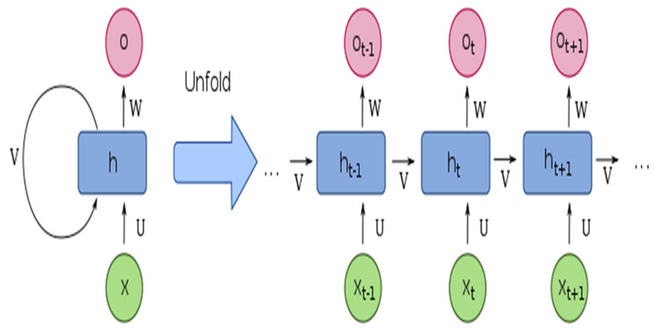


FIGURE 2 Folded and unfolded visualization of a recurrent neural network.

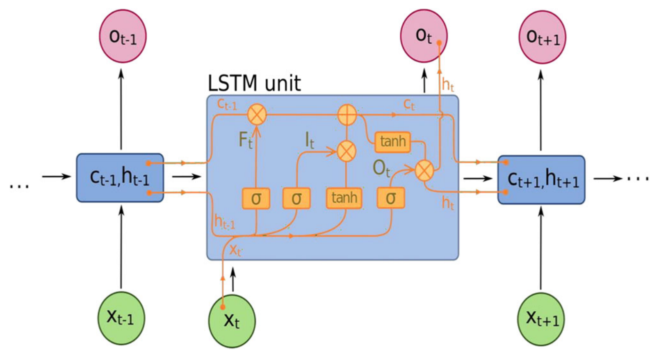


FIGURE 3 Visualization of long short-term memory unit.[17]

Additionally, a GRU uses the hidden state instead of the cell state to transmit information through the network.

LSTM and GRU are more flexible and better at capturing long-term dependencies than traditional RNNs. However, these networks are limited in their ability to retain past information, as the inputs only see information from the past. To address this, bidirectional LSTM (BiLSTM) (depicted in Figure 4) networks allow inputs to be fed through both the forward and backward layers, preserving information from both past and future at each time step. This provides the network with a better understanding of the context of the data, which can lead to improved predictions. Detailed theoretical explanations of LSTM, GRU, and BiLSTM can be found in [19, 20].

3.2 | RF theoretical descriptions

The RF algorithm is a popular ensemble of decision trees known for its simplicity and effectiveness. Although it can be applied to various data sets and deliver satisfactory results, it may not outperform specialized models in time series prediction and forecasting. A detailed explanation of RF can be found in [21, 22] but a brief overview will be provided in this section. The RF method

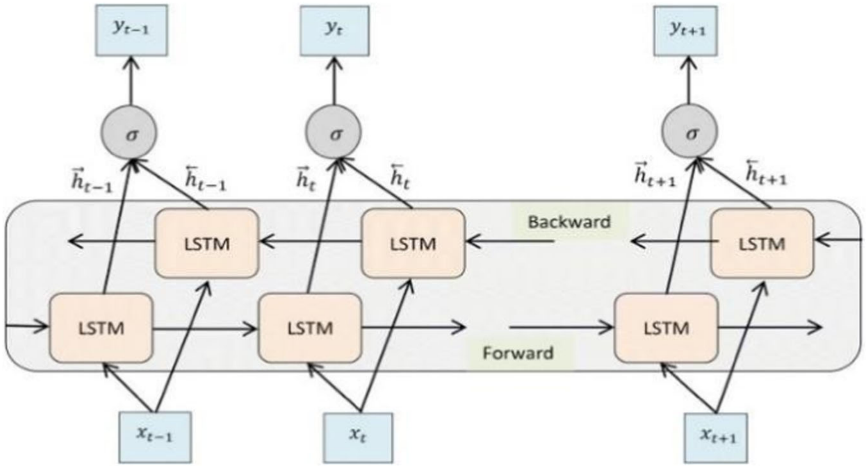


FIGURE 4 Visualization of bidirectional long short-term memory unit.[18]

combines numerous extremely randomized decision trees to obtain accurate and stable estimations. Each decision tree, known as a classification and regression tree, recursively divides the input space into small rectangular sections based on a predetermined split criterion and fits a simple model, typically a constant value, to each section. The RF strategy employs a group of weak learners to generate a strong learner by combining the decision trees, enhancing the algorithm's performance. To increase ensemble diversity, RF employs decision trees constructed from bootstrap samples of the original data set, thereby limiting classifiers to work on distinct random subsets of the complete feature space.[23] In this process, k bootstrap samples are randomly selected, and each sample is fitted using a regression tree. The average values of the k regression trees are then used for estimation. In this research, we employ the RF method to estimate RUL, generating bootstrap samples similarly to the bagging algorithm. This approach generates non-correlated trees from several training samples, providing noise immunity. However, the RF model selects random predictor variables at each split rather than fitting the tree with all the training data, improving accuracy and reducing the residual sum of squares to the greatest extent possible.[24]

3.3 | Experimental and software design setup

The proposed techniques will be comprehensively implemented using Python (version 3.6) along with popular data science-related packages, including NumPy and pandas, enabling efficient data navigation and manipulation. The RF algorithm will be implemented through the Scikit-learn library, while the neural networks (LSTM, BiLSTM, and GRU) will be constructed using TensorFlow 2.0 with the Keras module. These powerful libraries offer flexibility in fine-tuning and enable rapid model construction.

To optimize the performance of the models, hyperparameters will be fine-tuned using a random grid search to identify the most optimal combination. Leveraging TensorFlow and Keras allows for GPU-based computations, significantly reducing computation time compared to traditional CPU-based computations.

For evaluation, regression-oriented metrics such as RMSE, mean absolute error (MAE), MAPE, and symmetric mean absolute percentage error (sMAPE) will be utilized to assess model accuracy. The predictions will be compared to an existing method based on an ANN for performance comparison.

4 | DATA SET DESCRIPTION AND FILTRATION

The data set used is a publicly available vibration data set from a 2 MW wind turbine, particularly its high-speed shaft bearing driven by a 20-tooth pinion gear. The data set utilized in this study is sampled at 97,656 samples per second, capturing 6 s of vibration and tachometer data daily over a span of 50 days. This high-resolution data collection is critical for analyzing the nuanced behaviors of the high-speed shaft bearing in the wind turbine.[25]

A vibration results from an object moving or rotating back and forth from its normal stationary position. This cycle is captured every second via an accelerometer or a high-speed laser. Moreover, vibrations can be presented in multiple forms. The form present in our data set is that of acceleration (measured in g). It is worth to mentioned that while high-speed bearings can indeed be replaced individually, their failure can have significant implications for the overall performance and RUL of the wind turbine. When a high-speed bearing or gearbox fails unexpectedly, it can lead to unplanned downtime and costly repairs or replacements. The downtime directly affects the availability and productivity of the wind turbine, resulting in a loss of power generation and potential revenue. Therefore, special attention should be given to high-speed shaft bearing.

Initially, the data set format was a MATLAB. mat file, so it needed to be converted to a python friendly format (dictionaries and data frame) to further analyze. It should be noted that the proposal uses features as input while health indicator as output where all data in a day (time window) is used to compute the health indicators. It is important to note that the datasets used in this study contain both environmental and mechanical noise, which can obscure the true signals related to component health. To mitigate the impact of such noise, we employed advanced filtration techniques, including feature extraction and spectral kurtosis analysis. These methods are crucial for isolating fault-related features from the noise, thereby enhancing the accuracy of the RUL predictions.

Pre-processing of these data is essential because it ensures that the data is reliable, consistent, and in a suitable format for analysis and or modeling. It can also help improve the accuracy and efficiency of the proposed data driven algorithms and enable the identification of relevant features and patterns that may be useful in making informed decisions. We used[26] work to do pre-processing and feature engineering on the vibration data to produce 15 additional features in both the time and frequency domain as shown in Figure 5. The features were then ranked based on their monotonicity using the sign method (Equation 1).[27] A monotonic trend is when there's a consistent increase or decrease over time.

$$Monotonicity = \frac{1}{M}\sum_{j=1}^{M}\left|\sum_{k=1}^{N_{j-1}}\frac{s(x_j(k+1)-x_j(k))}{N_{j-1}}\right|. \tag{1}$$

Principal component analysis (PCA) is a popular technique used in data analysis and machine learning to reduce the dimensionality of the data while retaining most of its important features.

In the context of wind turbine RUL prediction, PCA can be used to reduce the number of features extracted from the vibration signals while retaining the most significant ones that contribute to the model's predictive power.

The selection of relevant features is crucial for the accuracy and effectiveness of machine learning algorithms. In the context of wind turbine RUL prediction, the selection of features extracted from the vibration signals of the high-speed shaft bearing is essential for accurately estimating the remaining useful life and predicting potential failures.

In this study, four features with high monotonicity were selected for PCA, as shown in Figure 5. Monotonicity refers to the degree of change in a feature over time and is an essential characteristic for selecting features that are relevant for RUL prediction. The four features selected exhibited high monotonicity, indicating that they are likely to be significant indicators of the health status of the high-speed shaft bearing.

PCA was then applied to the selected features to reduce the dimensionality of the feature space while retaining most of its essential information. The PCA reduced the features to two principal components, as displayed in Figure 6. The plot indicates a clear trend of increasing component values towards the failure point over time, suggesting that the extracted features are relevant for predicting the health indicator and RUL. The first PCA component was chosen as the target to predict the health indicator since it showed the most significant variation and contributed the most to the model's predictive power. The health indicator is a numerical value that represents the health status of the high-speed shaft bearing and is used as a measure of the remaining useful life. By predicting the health indicator, the model can estimate the remaining useful life of the high-speed shaft bearing and provide early warnings of potential failures, enabling timely maintenance and repair actions.

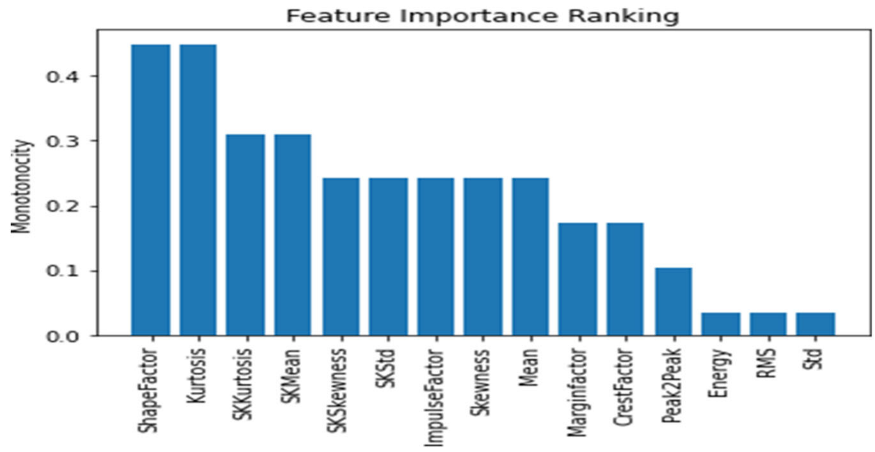
Overall, the selection of relevant features and the use of PCA in this study enabled the reduction of the dimensionality of the feature space while retaining most of the essential features that contribute to the model's predictive power. This approach can potentially improve the accuracy and efficiency of machine learning algorithms, particularly when dealing with high- dimensional datasets.

Since the data set is a time series, additional preprocessing was required to make it stationary. Stationarity means that the statistical properties of the data remain constant over time, including the mean and variance. Stationary data is easier to model and forecast. To achieve stationarity, the data was differenced, as shown in Equation 2. Differencing involves taking the difference between the values at time t and time t−1. The differenced values will be used for modeling and predictions. Equation (2) will be used to calculate the comparable error score using following,

$$Y_{diff} = Y_t - Y_{t-1}. \tag{2}$$

Furthermore, we would split our data set into a training and testing set at a 60:40 split. Our data set would then be scaled and transformed using a

FIGURE 5 Features Importance ranking.



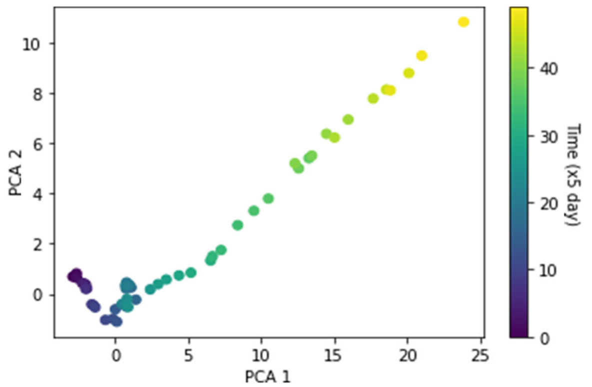


FIGURE 6 Plot of the first PCA against the second PCA.

MinMaxScaler which transforms our features by scaling them to a given range. For example, this range will be set to −1 and 1. The scaler will be fitted via the training set, and only then will it transform the training and test set. Scaling our data before feeding it through our neural networks can help mitigate problems with slow learning and convergence issues.

Health indicator is derived from the vibrational data to quantitatively represent the health of the bearing. The threshold for health indicator, beyond which the bearing is failing, is determined through historical failure data analysis. The assumption of stationarity was initially used for preliminary analytical simplicity. However, it is worth to note that health indicator varies with bearing conditions. The starting value of health indicator at 5.5 and ending at 22.5 were observed from the data set, indicating initial and final health states within the observed period.

5 | RESULTS AND DISCUSSION

The datasets were randomly selected, and the grid search hyperparameter optimization technique was used to select the best parameters for the LSTM model. The best hyperparameter values for the LSTM model were determined to be 50 epochs, 3 units, and a 0.2 dropout rate. For the GRU model, the best values were 50 epochs, 4 units, and a 0.3 dropout rate. For the BiLSTM model, the best values were 50 epochs, 3 units, and a 0.2 dropout rate. Finally, for the RF model, the best value was 250 estimators. These combinations led to lower RMSE scores and performed well when evaluated with other metrics.

In Figure 7, we plot the true values of the health indicator against the predictions made by the LSTM model on the remaining days from the test set. The LSTM model appears to be suitable for the task, as its predictions are close to the true values. The charts for the GRU and bidirectional LSTM models also look very similar, suggesting that these models perform similarly. However, the RF model, as shown in Figure 8, has predictions that deviate more compared to the neural networks.

To avoid overfitting in our neural networks, it is crucial to consider different methods. One approach is to use a simpler model by reducing the number of hidden layers and units. This is particularly advantageous, given that our data set is not especially large. Another approach is to add regularization to our model, such as incorporating a dropout layer. The dropout layer would randomly exclude a portion of the input units at each step of the training time, with the optimal dropout rate determined during the hyperparameter optimization stage. Since proposed algorithms are nondeterministic in the way the weights are initialized, we have run the fitting, prediction, and evaluation process 50 times for each method and for each of their hyperparameter combinations to reduce the impact of random initialization on the results.

Overall, proposed model results demonstrate that LSTM, GRU, and BiLSTM models can be effective in predicting the health indicator of a hydraulic system using a small set of features. We achieved the best results with the LSTM model, which had the lowest RMSE score. Our study highlights the importance of careful

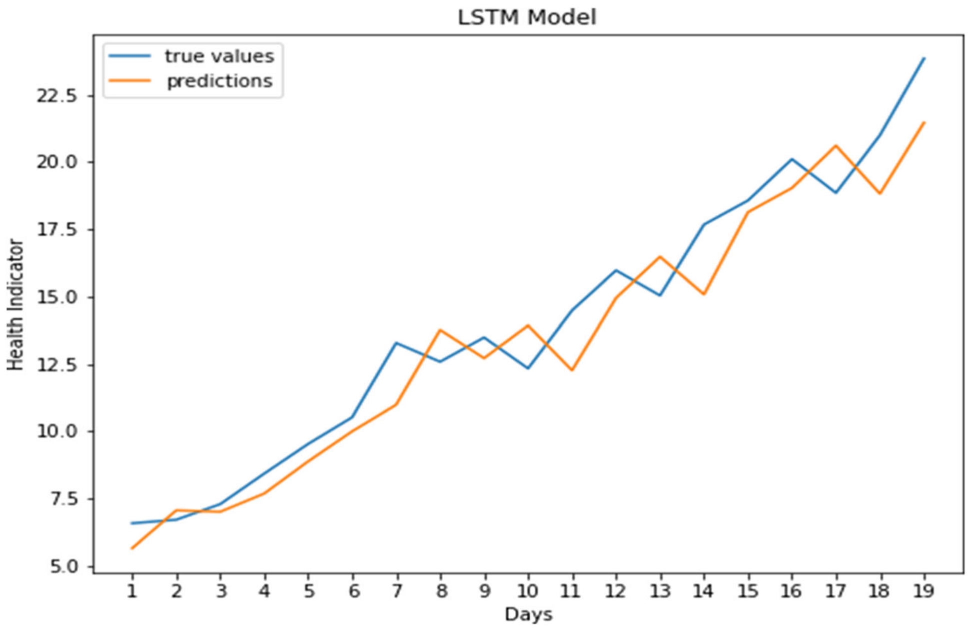


FIGURE 7 True values of the health indicator from the test set against the predicted values made by the LSTM model.

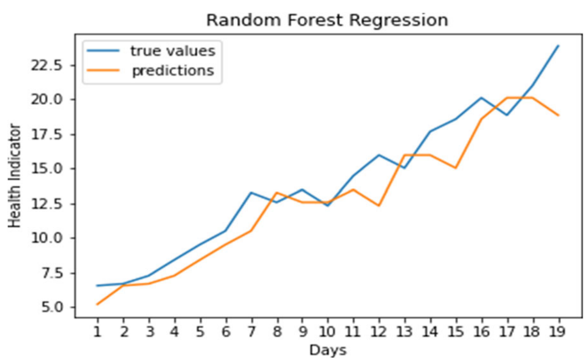


FIGURE 8 True values of the health indicator from the test set against the predicted values with Random Forest Regression.

hyperparameter tuning and regularization techniques to avoid overfitting and improve the performance of machine learning models.

Figure 9 illustrates the comparative performance of the LSTM and RF models in predicting the health indicator values of a wind turbine component over a 50-day period. The LSTM model, configured with 50 epochs, 3 units, and a 0.2 dropout rate, demonstrates superior predictive performance, closely aligning with the true values and exhibiting only minor deviations. This high accuracy can be attributed to the LSTM's capability to capture long-term dependencies in the time-series data, making it particularly effective for this task. In contrast, the RF model, with its 250 estimators, captures the overall trend but shows larger deviations and greater variability, indicating less stability and accuracy. The oscillations observed in the RF predictions highlight its limitation in handling the temporal dependencies inherent in the health indicator data. Despite its robustness and ability to handle diverse datasets, the RF model's performance is comparatively inferior to that of the LSTM, which effectively learns the underlying patterns in the data. This analysis underscores the LSTM model's potential for real-time health monitoring and predictive maintenance in wind turbines, enabling timely interventions and reducing downtime and maintenance costs. Meanwhile, the RF model, although less accurate, could still be useful in scenarios where quick, less precise estimates are acceptable, or computational resources are limited.

The RMSE, MAE, MAPE, and sMAPE are statistical performance metrics used in this work to evaluate the performance of data-driven RUL models. These performance metrics are stated mathematically as,

$$\mathrm{MAE}=\frac{\sum_{i=1}^{n}\mathrm{abs}(X_i{}'-X_i)}{n}, \tag{3}$$

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}\left(X_i'-X_i\right)^2}{n}}, \tag{4}$$

$$\mathrm{MAPE}=\frac{1}{n}\sum_{i}^{n}\left|\frac{X_i-X_i'}{X_i}\right|, \tag{5}$$

$$\mathrm{sMAPE}=\frac{1}{n}\sum_{i}^{n}\frac{\left|X_i-X_i'\right|}{\left(X_i+X_i'\right)/2}, \tag{6}$$

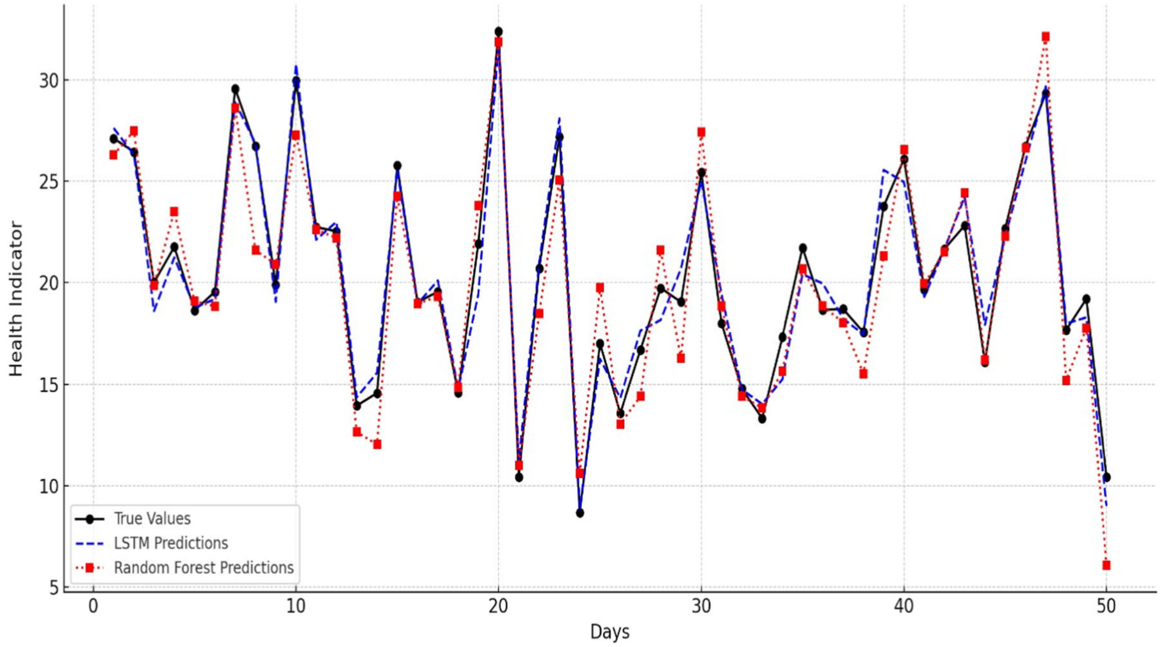


FIGURE 9 Comparative analysis of proposed models.

where X' are the predicated values for n different predictions, and X are the actual values. SSE is the sum of squared errors, and TSS, is the total sum of squares.

The square root of the mean of the squared difference between measured and anticipated power values is the RMSE. The MAE is the average of the absolute differences between measured and anticipated values. The MAPE is calculated by dividing the total number of absolute mistakes by the demand (each period separately). It is calculated as the average of percentage mistakes. Finally, sMAPE is a percentage (or relative) error-based accuracy measure. The calculated values of these metrics for RUL based models are tabulated in Table 1. We can see that the three neural networks all perform very similar, with the LSTM and BiLSTM variants slightly ahead. It is only the RF technique that noticeably falls behind. Our results show that engineered features from the time and frequency domain enable a health indicator to be formed that closely resembles the time to failure of a wind turbine's component. This health indicator can then be used as a prediction target in which more advanced variants of the RNN can be applied to. Our findings show that the tested LSTM, BiLSTM and GRU models performed very well once hyperparameters were optimized.

TABLE 1 Performance metrics results comparisons.

Methods	RMSE	MAE	MAPE	sMAPE
LSTM	1.3896	1.1855	8.4164	8.5830
BiLSTM	1.3885	1.1840	8.3942	8.5514
GRU	1.4144	1.2279	8.7260	8.9635
Random forest	1.9866	1.5481	10.7493	11.5631

Elasha et al.[28] applied supervised machine learning and deep learning methods to the same data set (used in this research) and achieved promising results. Their most comparable method to our findings is from their ANN, which performed the best and produced an RMSE of 2.3707. This specific ANN model is a classic feed-forward back- propagation model with one input and output layer and two hidden layers (nine and seven neurons in the first and second layer, respectively). A possible reason it didn't perform as well as our RNNs is that ANN, when compared to our three tested RNN variants, suffers more from the vanishing or exploding gradient problem. Therefore, the ANN would not be able to capture sequential information. Author of[29] and publish an open tool at MathWorks[25] where they used an exponential degradation model to predict the RUL of a wind turbine using the same data set, too. Their model would detect significant degradation in real-time and can update the model's parameter priors when a new observation becomes available. They, too, showed very promising results, achieving an RMSE of 1.4377.

A limitation of this study stems from the size of the data set. There are 50 recorded days, each only lasting 6 s. While this does a reasonable job, to get the full potential of an RNN, such as LSTM, it would have been more beneficial to get a more extensive data set spanning across more days. In addition, a larger data set would possibly have more hidden long-term dependencies within it that can be captured by the more advanced variants of the RNN. Still, proposed approaches can be easily applied to datasets much more extensive than what was used; only the hyperparameters and model complexity would need to be adjusted.

6 | CONCLUSION AND FUTURE WORK

This study introduces a robust data-driven approach for accurately predicting the RUL of a wind turbine's high-speed shaft bearing. Leveraging the potential of three variations of RNNs - LSTM, BiLSTM, and GRU - we successfully captured long-term dependencies in time series data and effectively mitigated the vanishing gradient problem. Additionally, to complement the RNNs' capabilities, we employed the versatile RF model, suitable for handling diverse datasets.

The data set used in this research was obtained from a 2 MW wind turbine, where field sensors continuously monitored and recorded raw vibration data from the high-speed shaft bearing. Through meticulous pre-processing and feature engineering, we enhanced the vibration signals, enabling the proposed models - LSTM, BiLSTM, GRU, and RF - to undergo training and make accurate predictions. Comparative evaluation of the models' performance revealed that LSTM and BiLSTM outperformed GRU and RF in predicting the health indicator and RUL of the wind turbine component.

This versatile methodology holds potential for extension to other components within wind turbines and other machinery, considering that vibration data can also be acquired from them. To further optimize the RNNs' performance, future work involves acquiring more extensive datasets and fine-tuning hyperparameters to fully exploit their capabilities. Moreover, exploring the Monte Carlo approach could facilitate the construction of confidence intervals, empowering operators to make informed maintenance decisions based on uncertainty quantification.

ACKNOWLEDGEMENTS

This research is supported by the Department for Science, Innovation & Technology (DSIT), UK under Tactical Fund Programme.

ORCID

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

REFERENCES

1. Department for Business, Energy & Industrial Strategy (2021), Digest of UK Energy Statistics (DUKES) 2021, Chapter 6: Renewable sources of energy. Retrieved from https://www.gov.uk/government/statistics/renewable-sources-of-energy-chapter-6-digest-of-united-kingdom-energy-statistics-dukes
2. IRENA report on Wind power. Available online at https://www.irena.org/costs/Power-Generation-Costs/Wind-Power. Accessed on December 15, 2021.
3. Berghout T, Mouss L-H, Bentrcia T, Benbouzid M. A semi-supervised deep transfer learning approach for rolling-element bearing remaining useful life prediction. *IEEE Trans Energy Convers*. 2022. 37(2):1200-1210. doi:10.1109/TEC.2021.3116423
4. Mishnaevsky, L, Thomsen K. Costs of repair of wind turbine blades: influence of technology aspects. *Wind Energy*. 2020;23(12):2247-2255.
5. Berghout T, Mouss LH, Kadri O, Saïdi L, Benbouzid M. Aircraft engines remaining useful life prediction with an adaptive denoising online sequential extreme learning machine. *Eng Appl Artif Intell*. 2020;96:103936.
6. Ding Y, Ding P, Jia M. A novel remaining useful life prediction method of rolling bearings based on deep transfer auto- encoder. *IEEE Trans Instrum Meas*. 2021;70:1-12.
7. Meddour I, Messekher SE, Younes R, Yallese MA. Se- lection of bearing health indicator by GRA for ANFIS-based forecasting of remaining useful life. *J Brazilian Soc Mech Sci Eng*. 2021;43(3):144.
8. Teng W, Han C, Hu Y, Cheng X, Song L, Liu Y. A robust model- based approach for bearing remaining useful life prognosis in wind turbines. *IEEE Access*. 2020;8:47133-47143.
9. Sanchez H, Escobet T, Puig V. Health-aware model predictive control of wind turbines using stifness degradation approach. *IFAC-PapersOnLine*. 2020;53(2):10348-10353.
10. Pan D, Liu JB, Cao J. Remaining useful life estimation using an inverse Gaussian degradation model. *Neurocomputing*. 2016;185:64-72.
11. Zhang J, Wang P, Yan R, Gao RX. Long short-term memory for machine remaining life prediction. *J Manuf Syst*. 2018;48:78-86.
12. Wang B, Lei Y, Li N, Li N. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. in *IEEE Trans Reliab*. 2020;69(no. 1):401-412. doi:10.1109/TR.2018.2882682
13. Ma M, Mao Z. Deep-convolution-based LSTM network for remaining useful life prediction. in *IEEE Trans Indus Inform*. 2021;17(no. 3):1658-1667.
14. Liu H, Song W, Niu Y, Zio E. A generalized Cauchy method for remaining useful life prediction of wind turbine gearboxes. *Mech Syst Signal Process*. 2021;153:107471. doi:10.1016/j.ymssp.2020.107471
15. A Beginner's Guide to Recurrent Networks and LSTMs. Available online at https://www.cnblogs.com/hansjorn/p/5398328.html
16. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(Number 8):1735-1780.
17. Article from wikipedia. Long Short-Term Memory. [Online image] CC BY-SA 4.0 Retrieved August 15, 2021, from https://commons.wikimedia.org/wiki/File:LongShort-TermMemory.svg
18. Huddar MG, Sannakki SS, Rajpurohit VS. Attention-based multimodal contextual fusion for sentiment and emotion classification using bidirectional LSTM. *Multimed Tools Appl*. 2021;80(9):13059-13076.
19. Cheng Y, Wu J, Zhu H, Or SW, Shao X. Remaining useful life prognosis based on ensemble long short-term memory neural network. in *IEEE Trans Instrument Measure*. 2021;70:1-12.

20. Choe D, Kim HC, Kim MH. Sequence-based modeling of deep learning with LSTM and GRU networks for structural damage detection of floating offshore wind turbine bladesrenew. *Energy*. 2021;174:218-235.
21. Pang J, Chen Y, He S, Qiu H, Wu C, Mao L. Classification of friction and wear state of wind turbine gearboxes using decision tree and random forest algorithms. *J Tribol*. 2021;143(9):091702.
22. Chen Y, Zheng W, Li W, Huang Y. Large group activity security risk assessment and risk early warning based on random forest algorithm. *Pattern Recognit Lett*. 2021;144:1-5.
23. Sammut, C, Webb, G. *'Encyclopedia of machine learning and data mining'*. Springer; 2017.
24. Breiman L, Friedman J, Olshen R, et al. *'Classification and regression trees'*. Chapman and Hall; 1984.
25. MathWorks. Wind Turbine High-Speed Bearing Prognosis. [Online] Retrieved June 10, 2021, https://uk.mathworks.com/help/predmaint/ug/wind-turbine-high-speedbearing-prognosis.html
26. Bechhoefer, E, Van Hecke, B, He, D. Processing for improved spectral analysis. *Ann Conf PHM Soc*. 2013;5(1). doi:10.36001/phmconf.2013.v5i1.2220
27. MathWorks, 2018. Quantify monotonic trend in condition indicators. [Online] Retrieved December 15, 2021, from https://uk.mathworks.com/help/predmaint/ref/monotonicity.html
28. Elasha F, Shanbr S, Li X, Mba D. Prognosis of a wind turbine gearbox bearing using supervised machine learning. *Sensors*. 2019;19(14):3092.
29. 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

How to cite this article: Pandit R, Santos M, Sierra-García JE. Comparative analysis of novel data-driven techniques for remaining useful life estimation of wind turbine high-speed shaft bearings. *Energy Sci Eng*. 2024;1-11. doi:10.1002/ese3.1911