

Comparative Analysis of Support Vector Machine, Random Forest and k-Nearest Neighbor Classifiers for Predicting Remaining Usage Life of Roller Bearings

Rajkumar Palaniappan*

College of Engineering, Department of Mechatronics Engineering, University of Technology Bahrain, Salmabad, Kingdom of Bahrain
r.palaniappan@utb.edu.bh

Keywords: RUL, SVM, KNN, random forest, machine Learning, bearing

Received: February 15, 2024

This research article aims to predict the remaining usage time of roller bearings using machine learning algorithms. The specific classifiers employed in this study are Support Vector Machines, Random Forest Classifier, and k-Nearest Neighbors. The predictive model takes into account various features including temperature, speed, load, dimensions of the inner and outer rings, width, vibration amplitude, vibration frequency, lubricant type, and lubricant viscosity. Data for training and testing the model were collected using a custom-made single bearing test rig. The target output variables are divided into intervals representing different percentages of remaining usage time. Principal component analysis (PCA) is utilized to identify the most influential features from the data. A ten-fold cross-validation method is employed for training and testing the classifiers. The features extracted through PCA are then fed into the classification model. The results show that the Support Vector Machines achieve the highest mean classification accuracy of 96.74%, followed by the Random Forest Classifier with 95.95%, and the k-Nearest Neighbors classifier with 91.77%. The study concludes that the Support Vector Machines outperform the Random Forest Classifier and k-Nearest Neighbors. Future research directions include exploring the application of deep learning algorithms to further enhance the predictive accuracy of the model. Additionally, conducting experiments with a larger and more diverse dataset, encompassing various operating conditions and types of bearings, would provide a broader understanding of the model's performance and generalizability.

Povzetek: Raziskava primerja SVM, Random Forest in k-Nearest Neighbor algoritme za napovedovanje preostale življenjske dobe valjčnih ležajev, pri čemer SVM doseže najvišjo točnost z 96,74%.

1 Introduction

Remaining Usage Life (RUL), also known as Remaining Useful Life, is a fundamental concept in the field of predictive maintenance. It involves estimating or predicting the amount of time that a specific component, system, or asset can be utilized before it fails or reaches a predefined threshold. RUL estimation is a critical aspect of maintenance planning and plays a significant role in optimizing maintenance strategies, minimizing downtime, reducing costs, and enhancing the overall reliability and performance of equipment (Lei et al., 2018; Nejjar et al., 2024; Palaniappan, Nataraj, Noaman, et al., 2023). The primary objective of predicting the RUL of a component or system is to proactively schedule maintenance actions based on its expected remaining operational life (Aberkane & Elarbi-Boudiher, 2022; Yaseen et al., 2022). By accurately estimating the RUL, maintenance activities can be planned and executed in advance, ensuring that necessary interventions are carried out before the occurrence of unexpected failures (Guo et al., 2017; Zhao et al., 2017). This approach helps prevent costly downtime, minimize the risk of catastrophic failures, and optimize the utilization of resources. Predicting the RUL

of a component or system requires an understanding of its degradation patterns and the ability to monitor its health condition (Ferreira & Gonçalves, 2022). Degradation patterns refer to the changes in the component's performance or health indicators over time, which are indicative of its remaining operational life (Fan et al., 2020; L. Zhang et al., 2018). These patterns can be characterized by analyzing historical data, observing trends in sensor readings, or studying the behavior of similar components under similar operating conditions. Monitoring the health condition of a component or system involves collecting relevant data through various sensors, such as temperature sensors, vibration sensors, acoustic sensors, or oil analysis sensors (Baptista et al., 2019; Palaniappan, Nataraj, Ismail, et al., 2023a; Yan et al., 2020). These sensors continuously measure key parameters that reflect the health and performance of the component. By monitoring these parameters, deviations from normal operating conditions can be detected, and the degradation process can be tracked. To estimate the RUL accurately, various data-driven techniques and algorithms are utilized (Chen et al., 2019; Elmahallawy et al., 2022). Machine learning algorithms, such as regression models, neural networks, decision trees, and support vector

machines, are commonly employed to analyze historical data, extract patterns, and predict the remaining operational life (Rathore & Harsha, 2022; Zhou et al., 2020). These algorithms utilize features or variables derived from the collected data to establish relationships between the degradation patterns and the expected RUL. The process of RUL estimation is not limited to a single approach or algorithm (Karim et al., 2023; Liu, 2024). It often involves iterative steps of data preprocessing, feature selection, model training, and validation. The quality and availability of data, the choice of appropriate degradation models, the selection of relevant features, and the consideration of uncertainties and variability are factors that can influence the accuracy of RUL predictions (Nabi et al., 2021). Advancements in technology, such as the Internet of Things (IoT) and edge computing, have enabled real-time RUL estimation (Khan et al., 2024; Lei et al., 2018; Palaniappan, Nataraj, Ismail, et al., 2023b). By integrating sensor devices with connectivity capabilities and leveraging edge computing resources, it is possible to continuously monitor the health condition of assets, collect real-time data, and provide up-to-date RUL predictions. Real-time RUL estimation allows for timely maintenance interventions, enables condition-based maintenance decisions, and enhances overall asset management efficiency (Nejjar et al., 2024; L. Zhang et al., 2018). In conclusion, Remaining Usage Life (RUL) estimation is a crucial concept in predictive maintenance. It involves predicting the remaining operational life of a component or system based on its degradation patterns and health condition monitoring. Accurate RUL estimation enables proactive maintenance planning, minimizes downtime, reduces costs associated with reactive maintenance, and enhances the overall reliability and performance of equipment. By utilizing data-driven techniques and algorithms, organizations can optimize maintenance strategies and ensure the effective utilization of resources.

2 Literature review

In recent years, there has been a lot of interest in applying machine learning approaches to predict machinery's remaining useful life (RUL) (Esfahani et al., 2021; Li et al., 2019). This literature review provides an overview of several studies and approaches to RUL prediction with machine learning. Regression models are a popular strategy for RUL prediction.

Baptista et al. investigated the use of the Kalman filter in data-driven prognostics, which includes a training stage for building a data-driven model and a prediction stage for estimating the end of life and remaining useful life of systems (Baptista et al., 2019). The Kalman filter is well-known for its integrated and resilient properties. The paper examines the performance of the Kalman filter in five data-driven prognostics systems that employ field data from an aeroplane bleed valve: neural networks, generalized linear models, k-nearest neighbors, random forests, and support vector machines. The results show

that Kalman-based models have superior precision and convergence. The Kalman filtering technique increases the accuracy and bias of the original regression models, especially when the equipment approaches its end of life. Among the approaches, the nearest neighbor's method demonstrated the greatest overall improvement, implying that Kalman filters may be particularly effective for instance-based methods.

Zhou et al. proposed a model for calculating the remaining useful life of battery cells in a pack. It uses k-nearest neighbor regression and includes information from all cells. A differential evolution technique optimizes the model's parameters. The approach determines the remaining useful life by averaging the useful lives of related cells. The approach yields an average error of 9 cycles, with the best estimation yielding an error of 2 cycles. Estimations are performed in 10 milliseconds. Accuracy improves as the number of tested and nearby cells increases. Compared to particle filter and support vector regression, the technique reduces estimation errors by 83.14% and 89.79%, respectively. The results support the method's usefulness in calculating the remaining usable life of lithium-ion cells (Zhou et al., 2020).

Yan et al., introduced a novel method for predicting the remaining useful life (RUL) of rolling element bearings, which are critical components in rotating machines (Yan et al., 2020). The method uses dimensionless data to assess bearing degradation and a hybrid degradation tracing model to estimate RUL optimally. Two novel metrics are proposed to reflect the vibration intensity of bearings, increasing sensitivity to incipient faults and decreasing variations. These metrics are used to identify the beginning of prediction and provide a dimensionless failure threshold. A support vector machine (SVM) classifier is used to determine the degradation stage with high accuracy, and it is trained using fitted measurements from a generalized degradation model. Five degradation stages have been defined for classification. However, actual measurements are used as inputs during the prediction process. The hybrid degradation tracing model uses the best RUL prediction to follow the deterioration progress based on classification findings. The suggested method is validated on public bearing datasets and compared to existing methods, confirming its effectiveness within a reasonable error range. Because the proposed measurements are dimensionless, this method can be used under a variety of operating situations.

In the work of Fan et al., proposed a transfer learning (TL) method based on feature representation for predicting the remaining useful life (RUL) of equipment (Fan et al., 2020). This approach solves the situation in which samples with previously unknown circumstances are encountered in the target domain, but labels are only available in the source domain. We use the Consensus Self-Organizing Models (COSMO) deviation detection method to create transferable attributes that capture each equipment's distinctiveness when compared to its peers. Our TL method's efficiency is proved utilizing the NASA

Turbofan Engine Degradation Simulation Data Set. Models with COSMO transportable features outperform other methods for predicting RUL, especially when the target domain is more complicated than the source domain.

Kang et al., presented a novel machine learning-based approach for automating the prediction of equipment failure in continuous production lines, with a focus on estimating the remaining useful life (RUL) (Kang et al., 2021). The suggested model includes normalization, principal component analysis for pre-processing, interpolation, grid search for parameter optimization, and the multilayer perceptron neural network (MLP) method. The approach is assessed using a case study of predicting engine RUL using NASA turbo engine information. The experimental results show that the suggested model is effective at predicting the RUL of turbo engines and greatly improved predictive maintenance outcomes.

Rathore & Harsha proposed a data-driven prognostics method for determining the remaining operational life of bearings (Rathore & Harsha, 2022). The method employs run-to-failure data from test rig tests to extract time-domain properties. Sudden changes in these traits signal the onset of defects that lead to failure. A monotonicity measure is used to choose the ideal feature set for

expressing bearing degradation. Dimension reduction and fusion are achieved using principal component analysis (PCA), which yields a unidimensional health indicator (HI). The oscillations in the HI are smoothed with a Weibull failure rate function (WFRF) and approximated using a nonlinear least-squares technique. By inverting the model, anticipated time values and the bearing's remaining operational life are estimated and compared to actual experimental results. Performance assessment measures including MAPE, MSE, RMSE, and bias are used. Furthermore, an online degradation state classification approach employing a k-nearest neighbor (KNN) classifier is built, resulting in good accuracy as seen by the ROC curve (receiver operating characteristic curve) with an AUC (Area under the ROC Curve) value of 0.94. Within 95% confidence levels, the predicted remaining useful life (RUL) closely approximates the actual RUL, with some variations. The model exhibits promising performance and can be used to estimate the remaining useful life of bearings.

A summary of the recent literature with equipment/component for which RUL is predicted using machine learning techniques along with dataset used, preprocessing, feature extraction, methodology and outcome are tabulated in Table 1.

Table1: Recent works related to RUL

Reference	Equipment/Component	Dataset	Preprocessing/Feature Extraction	Methodology	Outcome
(L. Zhang et al., 2018)	lithium-ion cell	ARBIN BT2000 is used for data collection	Differential evolution optimization	SVM, KNN and particle filter	KNN performed better compared to the other classifiers and the metric used is relative error.
(Baptista et al., 2019)	aircraft bleed valve	584 data recorded between 2010 to 2015	Kalman Filter	SVM, Neural Network, Random Forest, KNN, generalized linear models	All the classifiers with Kalman filter performed better compared to the classifiers without Kalman filter. Relative accuracy was used as metric.
(Yang et al., 2019)	Bearing dataset (acceleration and temperature data for 17 bearings)	PRONO STIA dataset	Cutting vibration signals into segments as training samples	Convolution Neural Network	Satisfactory results were obtained and the metric used was RMSE.
(Yan et al., 2020)	Bearing	force lubrication, constant speed and load - IMS dataset and	Relative Mean Square and Relative Mean Square (IRRMS)	SVM	98.88% Classification accrcy was obtained

		PRONO STIA dataset			
(Fan et al., 2020)	Turbofan Engine Degradation	NASA Turbofan Engine Degradation Simulation Data Set	transfer learning	Self organizing Maps	Root Mean Square Error – RMSE and mean absolute percentage error – MAPE
(Kang et al., 2021)	Turbofan Engine Degradation	NASA turbo engine datasets	Normalization and Principal component analysis.	Multilayer Perceptron (MLP)	Mean Square error
(Rathore & Harsha, 2022)	Roller bearing	Roller bearing dataset	Principal component analysis	KNN	MAPE, MSE, RMSE 95% confidence in predicting the RUL
(Y. Zhang & Zhao, 2023)	Lithium ion Battery	124 commercial lithium iron phosphate (LFP)/graphite cells	variance feature	Gaussian process regression	More than 85% classification accuracy reported
(Kumar & Upadhyaya, 2023)	rolling element bearing (REB)	Vibration signals, 10,000 data	continuous Wavelet transform (CWT) and extraction of statistical features from CWT coefficients.	KNN	83.3% Classification accuracy
(Motahari-Nezhad & Jafari, 2023)	Angular contact ball bearing	Acoustic emission signal	Prognostic feature selection	Multilayer Perceptron (MLP) and Radial Basis function (RBF) neural networks	MLP- Mean Squared error =7.86 RBF – Mean Squared error = 2.85
(Sharma et al., 2024)	axial piston pump	leakage volume	-	Auto-Regressive Integrated Moving-Average	RMSE Prediction of 28 months usage of axial piston pump before it fails was reported

Overall, machine learning techniques have proven to be effective in RUL prediction of various machinery components. The use of regression models, feature engineering, transfer learning, ensemble methods, deep learning, and uncertainty estimation are among the key approaches explored in the literature. These advancements contribute to enhanced maintenance strategies and improved operational efficiency in industrial settings. Existing studies in the literature have made significant contributions to bearing remaining usage life prediction. However, a common limitation among these studies is the limited scope of parameters considered for prediction. Most studies focus only on a subset of parameters, such as temperature, speed, load, or vibration frequency, while neglecting other influential factors. This limited parameter

set may not fully capture the complex dynamics of bearing behavior, leading to suboptimal accuracy in remaining usage life prediction. Additionally, most studies have used binary or discrete output variables, such as failure or degradation stages, rather than a continuous percentage-based measure of remaining usage life. Unlike previous studies that consider a limited set of parameters, this research incorporates a wide range of parameters that have been identified as influential in bearing performance. These parameters include temperature, speed, load, inner and outer ring diameter, width, vibration amplitude, lubricant type, and lubricant viscosity. By considering a more extensive parameter set, the study aims to capture a more accurate representation of the bearing's operational conditions and improve the prediction accuracy. While

most existing studies use binary or discrete output variables, this research focuses on predicting remaining usage life as a continuous percentage value. The use of a percentage-based output variable provides a more precise and informative measure of the remaining lifespan, enabling better decision-making in maintenance planning and resource allocation. The present study compares the performance of three different classifiers, namely Random Forest, K-Nearest Neighbors, and Support Vector Machine, in predicting bearing remaining usage life. By evaluating the results of multiple classifiers, the study aims to identify the most effective model for accurate remaining usage life prediction. This comparative analysis contributes to the selection of an optimal classifier for practical implementation in real-world scenarios. In summary, this research addresses the research gap in the existing literature on bearing remaining usage life prediction by considering a comprehensive set of parameters and utilizing a percentage-based output variable. The study's focus on multiple classifiers enables a comparative analysis to identify the most accurate prediction model. By filling this research gap, the findings of this study contribute to the advancement of predictive maintenance techniques, enhancing the reliability and efficiency of bearing performance assessment in various industrial applications.

3 Methodology

In this research article, a comprehensive multi-step methodology was employed to predict the remaining usage time of roller bearings using machine learning algorithms. The methodology, as depicted in Figure 1,

provided a systematic approach to address the prediction task. The dataset used in this study was collected from a custom-made single bearing test rig, specifically designed to capture various factors relevant to the bearing's operational conditions. These factors included temperature, speed, load, dimensions (such as inner and outer ring diameter, width), vibration amplitude and frequency, lubricant type, and viscosity. By encompassing these diverse features, the dataset aimed to capture the complexity of the bearing's behavior and its potential impact on the remaining usage time. To identify the most influential features within the dataset, Principal Component Analysis (PCA) was employed. PCA is a statistical technique that reduces the dimensionality of a dataset while preserving its essential variance. By applying PCA to the dataset, the researchers were able to extract the key features that significantly contributed to the prediction of the remaining usage time. This feature selection step facilitated a more focused analysis and enhanced the efficiency of the subsequent classification models. To ensure the reliability and generalizability of the predictive models, a ten-fold cross-validation method was employed. This technique involved dividing the dataset into ten subsets of approximately equal size. The models were then trained and evaluated ten times, with each subset serving as the testing set once. By applying cross-validation, potential biases and overfitting were mitigated, and the models' performance was robustly assessed. The mean classification accuracy, which measures the proportion of correctly classified instances, served as the evaluation metric for the models, providing a reliable indicator of their predictive capabilities.

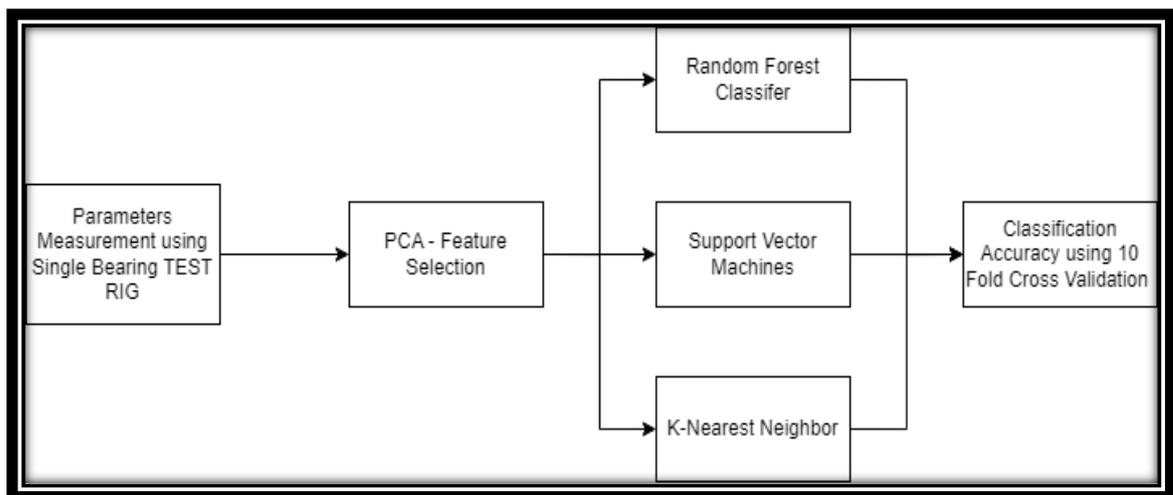


Figure 1: System block diagram

A) Data collection:

The bearing parameters utilized in this study were obtained from GDR's Tech Pvt Ltd, a company situated in Tamil Nadu, India. The data collection process involved

the use of a specifically designed single bearing test rig, which is depicted in Figure 2. This test rig was instrumental in gathering the necessary information for the research. The roller bearing was naturally degraded and

ran until failure to collect the required data. The parameters collected encompassed a range of relevant factors, including temperature, speed, load, inner ring diameter, outer ring diameter, width, vibration amplitude, vibration frequency, lubricant type, and lubricant viscosity. The selection of parameters for roller bearing Remaining Useful Life (RUL) prediction is crucial for developing an accurate and effective predictive model. The chosen parameters should capture the key factors that influence the degradation and failure of roller bearings.

Roller bearings are sensitive to temperature variations as high temperatures can accelerate wear, lubricant degradation, and material fatigue. Monitoring temperature helps identify potential overheating conditions that can lead to bearing failure.

The rotational speed of roller bearings affects their operating conditions and influences factors such as lubrication effectiveness, load distribution, and contact stress. Higher speeds can lead to increased wear and fatigue.

The load applied to a roller bearing affects its stress levels, fatigue life, and overall performance. Excessive loading or variations in load can accelerate wear, leading to premature failure.

The dimensions of the inner and outer rings determine the bearing's load-carrying capacity and its ability to withstand external forces. Changes in ring diameter can affect the distribution of load and contribute to bearing degradation.

The width of a roller bearing impacts its load-carrying capacity and stiffness. A wider bearing generally has a higher load capacity, but it can also affect other factors such as lubrication and heat dissipation.

Vibration analysis provides insights into the condition of roller bearings. Abnormal vibration patterns can indicate faults such as misalignment, imbalance, or bearing defects. Monitoring vibration amplitude and frequency helps detect early signs of degradation.

The type of lubricant used in roller bearings significantly affects their performance and longevity. Different lubricants have varying properties, such as viscosity and temperature range, which impact the bearing's ability to reduce friction and wear.

Viscosity is a critical parameter that determines the lubricant's ability to maintain a protective film between rolling elements and raceways. Proper viscosity ensures efficient lubrication and reduces friction and wear.

By considering these parameters in roller bearing RUL prediction, you cover various aspects that influence the bearing's degradation and failure. This comprehensive approach allows for a more accurate assessment of the bearing's health and improves the effectiveness of maintenance strategies, such as condition-based maintenance or predictive maintenance, to optimize the bearing's lifespan and minimize unexpected failures. Multichannel data acquisition system was utilized to collect the data from the sensors. These parameters were

carefully chosen to encompass various aspects that could affect the remaining useful life (RUL) of the bearings. The collected database consisted of a total of 6180 data points. Each data point represented a specific RUL category. To be specific, there were 1000 data points representing a 5% RUL, 1000 data points representing a 10% RUL, 1000 data points representing a 15% RUL, 1000 data points representing a 20% RUL, and 2180 data points representing a RUL greater than 20%. This distribution of data points allowed for a comprehensive analysis of the bearing's behavior at different stages of its life cycle. By considering a diverse range of RUL values, the researchers aimed to develop a robust and accurate predictive model for estimating the remaining usage time of roller bearings. The raw data was found to be statistically insignificant using ANOVA method and hence application of machine learning is mandatory in classifying the data.

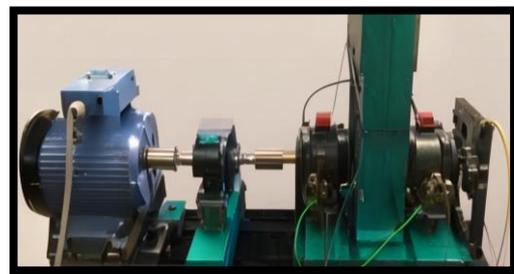


Figure 2: Data collection setup

B) Feature selection:

In the process of predicting the remaining usage time of roller bearings using Random Forest (RF), Support Vector Machines (SVM), and k-Nearest Neighbors (KNN) algorithms, Principal Component Analysis (PCA) plays a crucial role in feature selection. The selected features for PCA encompass a comprehensive set, including temperature, speed, load, inner ring diameter, outer ring diameter, width, vibration amplitude, vibration frequency, lubricant type, and lubricant viscosity. PCA enables dimensionality reduction by transforming the correlated variables into uncorrelated principal components. This transformation allows for a more concise representation of the data while preserving its essential variance. By analyzing the eigenvalues and loadings associated with the principal components, we gain insights into their contributions to the overall variability of the dataset. This information allows us to identify the most influential features, rank them according to their significance, and select the most informative ones for accurate predictions. The utilization of PCA for feature selection enhances the performance of the Random Forest Classifier and KNN algorithms by reducing dimensionality and focusing on the key features that have a substantial impact on the remaining usage time of roller bearings. By eliminating redundant or less informative features, PCA allows the algorithms to concentrate on the most critical aspects that drive the bearing's aging process. This approach not only improves the efficiency of the algorithms but also

provides a more interpretable and concise set of features, enabling better understanding and insights into the factors affecting the remaining life usage of roller bearings.

C) Model development:

The section explains the implementation of the Random Forest and k-Nearest Neighbors classifiers. It details the training process using ten-fold cross-validation, ensuring the models are robust and capable of generalizing well to new data.

Random forest classifier

The Random Forest Classifier is an ensemble learning algorithm widely used in machine learning for classification tasks. It constructs multiple decision trees during the training phase and combines their predictions to make final predictions. Each decision tree is built using a random subset of the training data and a random subset of features, which helps reduce overfitting and enhance the model's generalization ability. This randomness also enables Random Forests to handle high-dimensional feature spaces effectively and makes them robust to outliers and noise in the data. During prediction, each decision tree independently classifies the input data, and the class with the majority vote across all trees is selected as the final prediction. Random Forests can capture complex decision boundaries and capture intricate relationships between features. One of the advantages of Random Forests is their ability to estimate feature importance. By analyzing how much each feature contributes to the classification task, insights can be gained into the underlying relationships and factors driving the predictions. Additionally, Random Forests offer a good balance between bias and variance, which helps prevent overfitting and improves the model's generalization performance. However, it is important to note that Random Forests may not perform as well on datasets with severe class imbalance, and they can be computationally more expensive compared to simpler models like decision trees. Furthermore, the interpretability of Random Forests can be limited due to the ensemble nature of the model. Despite these considerations, the Random Forest Classifier remains a powerful and versatile algorithm widely used in various domains for its robustness, accuracy, and ability to handle complex classification tasks.

Support vector machines

The Support Vector Machine (SVM) classifier is a powerful algorithm for solving classification problems. SVM works by finding an optimal hyperplane that separates the data points of different classes in a high-dimensional feature space. The key idea behind SVM is to maximize the margin, which is the distance between the decision boundary and the nearest data points of each class. By maximizing the margin, SVM aims to find a

hyperplane that generalizes well to unseen data. SVM can handle both linearly separable and non-linearly separable data by using the kernel trick. The kernel function transforms the input features into a higher-dimensional space where the data becomes linearly separable. This allows SVM to capture complex decision boundaries and make accurate predictions. SVM also introduces a regularization parameter (C) to balance the trade-off between achieving a low training error and a low margin. By adjusting C , the user can control the flexibility of the model and prevent overfitting or underfitting. One of the advantages of SVM is its ability to handle high-dimensional data effectively. SVM constructs a decision boundary using a subset of training data points called support vectors, which are the data points closest to the decision boundary. This property makes SVM memory-efficient and suitable for datasets with a large number of features. Additionally, SVM is robust to noise and outliers in the data. The use of the margin ensures that SVM focuses on the most informative data points near the decision boundary, rather than being influenced by outliers. SVM has demonstrated strong performance in various domains, including image classification, text categorization, and bioinformatics. However, SVM's training time can be relatively high for large datasets, and the selection of appropriate hyperparameters, such as the choice of kernel and C , requires careful tuning. Nonetheless, SVM remains a popular and widely used classifier due to its versatility and ability to handle complex classification tasks.

k-Nearest Neighbors

The k-Nearest Neighbors (k-NN) algorithm is a non-parametric machine learning algorithm used for classification and regression tasks. In k-NN, the prediction for a new data point is determined by the majority vote or average of the values of its k nearest neighbors in the feature space. The algorithm does not require a training phase as it directly uses the training data for predictions. It is a flexible algorithm that can handle complex decision boundaries and capture non-linear relationships in the data. One advantage of k-NN is its simplicity and ease of implementation. However, there are considerations when using k-NN. The choice of the parameter k is critical, as a small k value may lead to overfitting, while a large k value may result in oversimplification. Additionally, k-NN can be sensitive to the scale of features, and data normalization is often necessary. As the number of data points increases, the computational cost of k-NN can also become a limitation. In classification tasks, k-NN is commonly used when the decision boundaries are not well-defined or when the data is not linearly separable. It can be particularly effective when dealing with multi-class classification problems. In regression tasks, k-NN can provide accurate estimates by averaging the values of its k nearest neighbors. However, it's important to note that k-NN suffers from the curse of dimensionality, where the

algorithm's performance can deteriorate as the number of features increases. Despite its limitations, k-NN remains a popular and versatile algorithm that is widely used in various domains, especially when interpretability and flexibility are important considerations.

Performance evaluation:

10-fold cross-validation is a widely used and reliable technique in machine learning for evaluating the performance and generalization ability of a model. It involves dividing the dataset into ten equal-sized subsets or "folds" and repeatedly training and testing the model on different combinations of training and validation sets. This approach provides a more robust estimate of the model's performance by reducing the impact of a single train-test split and considering the variability in performance across multiple iterations. By ensuring that each data point participates in both training and validation, cross-validation offers a comprehensive evaluation of the model's effectiveness and its ability to generalize to unseen data. It helps address issues related to data partitioning and sample bias, making it particularly valuable in scenarios with limited or imbalanced datasets. Additionally, cross-validation enables researchers to assess the stability and reliability of the model by observing performance variations across different folds. While 10-fold cross-validation is a widely adopted approach, other variations of cross-validation, such as stratified k-fold or leave-one-out cross-validation, may be more appropriate depending on the specific characteristics of the dataset and research objectives. These variations cater to scenarios with imbalanced datasets or when dealing with limited data. The choice of cross-validation technique should align with the requirements of the application or research study. Regardless of the specific approach chosen, cross-validation serves as a valuable

tool for estimating the performance of machine learning models and providing a more reliable evaluation of their generalization capabilities.

4 Results and discussion

This section presents the outcomes of a comparative analysis study that focuses on evaluating the performance of three machine learning algorithms - Support Vector Machine (SVM), Random Forest Classifier (RFC), and k-Nearest Neighbor (KNN) - for predicting the remaining usage life of roller bearings. The primary metric used to assess the performance of these algorithms is classification accuracy, which measures the proportion of correctly classified instances in the dataset. The analysis is conducted using a ten-fold cross-validation approach to ensure robustness and reliability in the evaluation process. Equation 1 is used in obtaining the mean classification accuracy of each classifier for each fold. Equation 2 is used to determine the sensitivity and Equation 3 is used to determine the specificity.

$$Accuracy = \left(\frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \right) * 100 \tag{1}$$

$$Sensitivity = TP / (TP + FN) \tag{2}$$

$$Specificity = TN / (TN + FP) \tag{3}$$

where,

TP- True Positive

TN- True Negative

FN- False Negative

FP – False Positive

Table 2: Classification accuracy for random forest classifier in classifying remaining usage life

Accuracy in %	n_ = 50 d = 15 s = 10 l = 2 f = 'sqrt'	n_ = 50 d = 10 s = 2 l = 1 f = 'sqrt'	n_ = 100 d = 10 s = 2 l = 1 f = 'sqrt'	n_ = 100 d = 10 s = 2 l = 1 f = 'sqrt'	n_ = 100 d = 10 s = 4 l = 1 f = 'sqrt'	n_ = 200 d = 10 s = 2 l = 1 f = 'sqrt'	n_ = 200 d = 15 s = 4 l = 1 f = 'sqrt'
Fold 1	89.45	94.40	94.69	96.83	91.92	89.68	88.60
Fold 2	87.41	94.40	94.21	95.55	91.84	89.03	88.86
Fold 3	86.89	93.39	94.83	96.46	93.03	88.86	88.35
Fold 4	85.63	93.89	94.89	95.88	92.17	89.85	88.69
Fold 5	86.12	93.95	93.69	96.85	92.40	88.61	88.59
Fold 6	87.35	94.59	94.45	95.87	92.98	89.09	87.42
Fold 7	86.93	94.98	94.34	94.38	91.11	89.40	88.29
Fold 8	89.45	94.96	95.56	96.10	93.45	89.91	87.08
Fold 9	87.68	92.65	94.31	95.78	91.37	89.92	88.02
Fold 10	89.31	93.10	93.55	95.75	92.44	89.08	88.53
Mean Accuracy	87.62	94.03	94.45	95.95	92.27	89.34	88.24

Where Number of Trees (n), Maximum Depth (d), Minimum Samples Split (s), Minimum Samples Leaf (l) Maximum Features (f) = 'sqrt' (square root of the total number of features)

Table 2 shows the results obtained using random forest classifier by tuning the following parameters: Number of Trees (n), Maximum Depth (d), Minimum Samples Split (s), Minimum Samples Leaf (l) Maximum Features (f) = 'sqrt' (square root of the total number of features). Across all parameter settings, the Random Forest classifier demonstrates relatively high accuracy in classifying the Remaining Usage Life, with mean accuracies ranging from 87.62% to 95.95%. The highest mean accuracy is achieved with the parameter settings: n = 100 (Number of Trees), d = 10 (Maximum Depth), s = 2 (Minimum Samples Split), l = 1 (Minimum Samples Leaf), and f = 'sqrt' (Maximum Features). The lowest mean accuracy is observed with the parameter settings: n = 50 (Number of Trees), d = 15 (Maximum Depth), s = 10 (Minimum Samples Split), l = 2 (Minimum Samples Leaf), and f =

'sqrt' (Maximum Features). It is worth noting that the parameter settings with higher values for the number of trees (n) tend to yield slightly better performance, as indicated by the higher mean accuracy values. Additionally, the parameter settings with lower maximum depth (d), lower minimum samples split (s), and lower minimum samples leaf (l) values tend to result in better performance. The choice of the maximum features (f) parameter as 'sqrt' (square root of the total number of features) appears to be effective in achieving good accuracy values, as seen across all parameter settings.

The sensitivity and specificity for the parameters which achieved highest mean classification accuracy for Random Forest classifiers are 90.56% and 97.14% respectively.

Table 3 : Classification accuracy for K-nn classifier in classifying remaining usage life

Accuracy in %	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10
Fold 1	78.11	84.74	89.97	92.63	85.56	84.48	82.26	78.50	78.64	78.71
Fold 2	77.53	85.52	88.28	91.44	85.94	84.02	82.47	76.24	75.03	78.90
Fold 3	80.73	85.43	88.54	92.48	86.48	84.69	80.09	77.14	77.15	77.55
Fold 4	81.24	84.03	90.39	91.53	86.62	83.51	82.74	78.31	78.55	75.42
Fold 5	81.57	84.26	89.61	92.32	85.04	84.21	80.25	76.74	77.53	77.95
Fold 6	80.80	84.03	90.77	92.57	84.93	84.83	81.25	76.39	77.24	76.61
Fold 7	82.43	85.94	90.23	91.63	85.04	83.52	80.83	76.81	78.14	75.94
Fold 8	81.00	84.81	90.14	90.52	84.94	83.57	82.85	78.60	76.74	75.24
Fold 9	82.25	85.71	90.75	91.57	86.66	83.86	82.97	78.54	75.94	78.72
Fold 10	83.41	84.12	89.97	91.00	86.88	84.04	80.82	78.68	77.15	77.80
Mean Accuracy	80.91	84.86	89.87	91.77	85.81	84.07	81.65	77.60	77.21	77.28

The table 3 represents the classification accuracy of a K-nearest neighbors (K-NN) classifier for classifying the Remaining Usage Life. The accuracy values are expressed in percentages, and the different values of K range from 1 to 10. The table also includes the accuracy values for each fold in a cross-validation process, as well as the mean accuracy across all folds for each value of K. On average, the highest accuracy values are achieved around K=4, where the accuracy ranges from 91.77% to 92.63%. The accuracy tends to decrease or stabilize for larger values of K beyond the optimal range of K=4 to K=5. The mean accuracy values range from 77.21% to 84.86%, showing the overall performance of the classifier across all folds and values of K.

The sensitivity and specificity for the parameters which achieved highest mean classification accuracy for KNN classifiers are 86.13% and 94.28% respectively.

The table 4 represents the classification accuracy of a Support Vector Machine (SVM) classifier for classifying the Remaining Usage Life. The accuracy values are

expressed in percentages. The table includes different combinations of hyperparameters for the SVM classifier, such as different values of C and gamma, and their corresponding accuracy values for each fold in a cross-validation process. The mean accuracy across all folds for each combination of hyperparameters is also provided. The accuracy of the SVM classifier varies depending on the combination of hyperparameters. The highest mean accuracy is achieved with Combination 4, where C=1, RBF kernel, and Gamma=0.001, with a mean accuracy of 96.75%. Generally, increasing the value of C or selecting a smaller Gamma tends to improve the accuracy of the SVM classifier. There is some variability in accuracy across different folds, indicating the potential impact of data partitioning on model performance.

The sensitivity and specificity for the parameters which achieved highest mean classification accuracy for SVM classifiers are 94.73% and 98.44% respectively.

Table 4: Classification accuracy for SVM classifier in classifying remaining usage life

Accuracy in %	C=0.01 RBF kernel Gamma =0.01	C=0.01 RBF kernel Gamma =0.001	C=1 RBF kernel Gamma =0.01	C=1 RBF kernel Gamma =0.001	C=10 RBF kernel Gamma =0.01	C=10 RBF kernel Gamma =0.001	C=25 RBF kernel Gamma =0.01	C=25 RBF kernel Gamma =0.001
Fold 1	91.33	91.19	95.60	97.18	95.96	93.34	92.24	92.14
Fold 2	91.45	92.20	95.17	95.96	94.98	94.63	93.72	94.02
Fold 3	89.54	89.24	94.00	95.61	94.95	94.95	92.81	94.85
Fold 4	88.27	90.76	93.29	98.74	93.99	95.47	94.86	93.48
Fold 5	89.03	91.58	96.47	97.27	95.10	94.01	94.90	92.23
Fold 6	90.14	92.21	93.63	95.99	93.13	94.75	92.02	93.98
Fold 7	89.29	91.83	96.50	95.77	93.23	95.32	92.69	92.43
Fold 8	88.25	91.96	94.42	95.62	95.68	94.25	94.80	91.34
Fold 9	90.36	92.15	95.77	98.75	95.80	94.60	94.34	94.38
Fold 10	88.13	89.29	96.95	96.64	95.27	93.51	94.71	93.97
Mean Accuracy	89.58	91.24	95.18	96.75	94.81	94.48	93.71	93.28

The Random Forest classifier achieves accuracy values ranging from 78.11% to 92.63%. The accuracy tends to increase initially with K (number of nearest neighbors) and reaches its peak around K=4. After K=4, the accuracy either decreases or remains relatively stable. The mean accuracy across all folds ranges from 77.21% to 84.86%. The SVM classifier achieves accuracy values ranging from 88.13% to 98.75%. Different combinations of hyperparameters, such as C (regularization parameter) and gamma, lead to varying accuracies. The mean accuracy across all folds ranges from 89.58% to 96.75%. The highest mean accuracy is achieved with Combination 4, where C=1, RBF kernel, and Gamma=0.001. The K-NN classifier achieves accuracy values ranging from 78.11% to 92.63%. The accuracy tends to increase initially with K and reaches its peak around K=4. After K=4, the accuracy either decreases or remains relatively stable. The mean accuracy across all folds ranges from 77.21% to 84.86%. The SVM classifier achieves the highest accuracy values among the three classifiers, with a range of 88.13% to 98.75%. The accuracy range for Random Forest and K-

NN classifiers is similar, ranging from 78.11% to 92.63%. The SVM classifier also exhibits the highest mean accuracy across all folds, ranging from 89.58% to 96.75%. The mean accuracy for Random Forest and K-NN classifiers ranges from 77.21% to 84.86%. The SVM classifier's performance is highly dependent on the choice of hyperparameters, such as C and gamma. Tuning these hyperparameters can significantly impact the accuracy. Both the Random Forest and K-NN classifiers show a similar pattern in terms of the optimal value of K. The accuracy tends to increase with K initially and reaches a peak around K=4, after which it either decreases or stabilizes. Overall, the SVM classifier demonstrates the highest accuracy and mean accuracy among the three classifiers, indicating its potential as a reliable classifier for classifying the Remaining Usage Life. However, it's important to note that the optimal choice of classifier depends on the specific dataset and classification task, and further analysis and experimentation may be needed to determine the most suitable classifier for a given scenario.

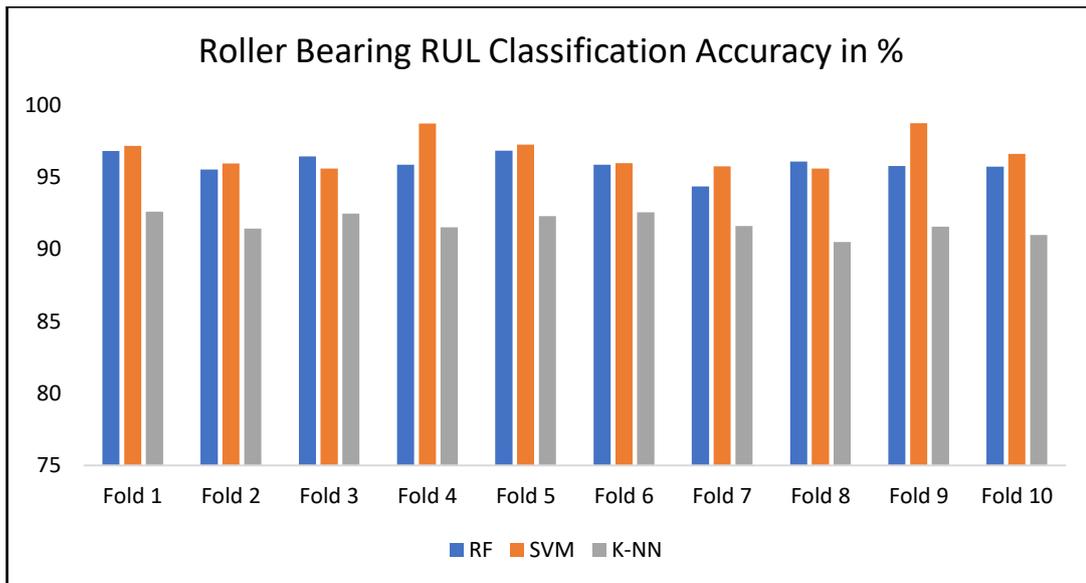


Figure 3: Roller bearing RUL classification accuracy in %

Figure 3 shows the graph for roller bearing RUL maximum classification accuracy for each classifier extracted from Table 1, Table 2 and Table 3. The Graph clearly shows that SVM classifier has the maximum classification accuracy and Random Forest classifier accuracy is similar to the SVM classifier. The classification accuracy obtained using KNN classifier is less compared to the other two classifiers.

The obtained results in this article cannot be directly compared with existing research in the literature due to the unique nature of the dataset used. Unlike most existing studies that utilize fewer than three parameters for predicting the remaining usage time of roller bearings, this research incorporates a comprehensive set of parameters including temperature, speed, load, inner and outer ring diameters, width, vibration amplitude, vibration frequency, lubricant type, and lubricant viscosity. This difference in dataset composition hinders a direct comparison with previous literature. The incorporation of a wide range of parameters in this study reflects a more realistic and holistic approach to bearing health monitoring and remaining useful life prediction. By considering multiple parameters, the model developed in this research captures the complexity and interdependencies of various factors that affect bearing degradation and remaining usage time. However, this also means that the results obtained cannot be directly benchmarked against previous studies that focus on a limited number of parameters.

Firstly, let's compare the component or equipment for which RUL prediction was carried out. From Table 1, it was observed that the work of (Yang et al., 2019) (Yan et al., 2020) (Rathore & Harsha, 2022) (Kumar & Upadhyaya, 2023) (Motahari-Nezhad & Jafari, 2023) also have developed RUL prediction for bearing which is like the present study which is used for predicting the RUL for

roller bearing. Secondly when we compare the dataset, most of the works use vibration signal/acoustic emission in predicting the RUL using machine learning algorithm. However, the present study uses more than one parameter in prediction of RUL. Next When we compare the preprocessing /Feature extraction method used, (Rathore & Harsha, 2022) used PCA and obtained 95% accuracy which is like the method used in this research. While comparing the classifier used, most of the researchers have used KNN, SVM and Neural networks. The outcome in terms of predicting the Rul have been satisfactory in most cases. While considering the number of parameters used for developing the model, the present study has performed better compared to the previous research.

While the lack of comparative literature limits the ability to assess the performance of the developed model in relation to existing approaches, it opens opportunities for future research. With the availability of more diverse and extensive datasets, a deep learning model could be developed to further improve the classification accuracy for predicting remaining usage time. Deep learning models, such as convolutional neural networks or recurrent neural networks, have demonstrated their effectiveness in handling complex datasets and capturing intricate patterns. By leveraging the power of deep learning and utilizing larger datasets encompassing various operating conditions and bearing types, it is possible to enhance the accuracy and generalizability of the predictive model. In conclusion, the uniqueness of the dataset used in this study, incorporating multiple parameters that are not typically found in existing literature, restricts the direct comparison of results. However, this limitation also highlights the potential for future research to explore more comprehensive approaches, such as deep learning, using larger and diverse datasets. By addressing these limitations and

further advancing the understanding of bearing health monitoring, we can enhance the accuracy and applicability of predictive models for remaining usage time estimation in practical industrial applications.

6 Conclusion

In conclusion, this research article successfully developed a predictive model for estimating the remaining usage time of roller bearings using machine learning algorithms. The study employed Support Vector Machines, Random Forest Classifier, and k-Nearest Neighbors as the specific classifiers. These algorithms effectively utilized various features, including temperature, speed, load, dimensions of the inner and outer rings, width, vibration amplitude, vibration frequency, lubricant type, and lubricant viscosity, to make accurate predictions. To train and evaluate the model, a custom-made single bearing test rig was utilized to collect data. The target output variables were segmented into intervals representing different percentages of remaining usage time, allowing for more precise predictions. Principal component analysis (PCA) was applied to identify the most influential features from the dataset, enhancing the model's performance. A ten-fold cross-validation method was employed to ensure robust training and testing of the classifiers. The results demonstrated that the Support Vector Machines achieved the highest mean classification accuracy of 96.74%, followed by the Random Forest Classifier with 95.95%, and the k-Nearest Neighbors classifier with 91.77%. These findings indicate that Support Vector Machines outperformed the other two algorithms in accurately predicting the remaining usage time of roller bearings. The study suggests several future research directions to further improve the predictive accuracy of the model. One potential avenue is exploring the application of deep learning algorithms, which have shown promising results in various domains. Additionally, conducting experiments with a larger and more diverse dataset, encompassing different operating conditions and types of bearings, would provide a more comprehensive understanding of the model's performance and its ability to generalize to real-world scenarios. Overall, this research contributes to the field of predictive maintenance by demonstrating the efficacy of machine learning algorithms in estimating the remaining usage time of roller bearings. The findings have practical implications for industries relying on these bearings, enabling them to optimize maintenance schedules and reduce unexpected failures.

Acknowledgement

The Author would like to acknowledge University of Technology Bahrain for providing resources and support. Additionally, the author expresses gratitude to GDR's Tech Pvt, Ltd for generously sharing the dataset that was

instrumental in the development and evaluation of our predictive model.

References

- [1] Aberkane, S., & Elarbi-Boudihir, M. (2022). Deep Reinforcement Learning-based anomaly detection for Video Surveillance. *Informatica*, 46(2). <https://doi.org/10.31449/inf.v46i2.3603>
- [2] Baptista, M., Henriques, E. M. P., de Medeiros, I. P., Malere, J. P., Nascimento, C. L., & Prendinger, H. (2019). Remaining useful life estimation in aeronautics: Combining data-driven and Kalman filtering. *Reliability Engineering & System Safety*, 184, 228–239. <https://doi.org/10.1016/j.res.2018.01.017>
- [3] Chen, Z., Li, Y., Xia, T., & Pan, E. (2019). Hidden Markov model with auto-correlated observations for remaining useful life prediction and optimal maintenance policy. *Reliability Engineering & System Safety*, 184, 123–136. <https://doi.org/https://doi.org/10.1016/j.res.2017.09.002>
- [4] Elmahallawy, M., Elfouly, T., Alouani, A., & Massoud, A. M. (2022). A Comprehensive Review of Lithium-Ion Batteries Modeling, and State of Health and Remaining Useful Lifetime Prediction. *IEEE Access*, 10, 119040–119070. <https://doi.org/10.1109/ACCESS.2022.3221137>
- [5] Esfahani, Z., Salahshoor, K., Farsi, B., & Eicker, U. (2021). A New Hybrid Model for RUL Prediction through Machine Learning. *Journal of Failure Analysis and Prevention*, 21(5), 1596–1604. <https://doi.org/10.1007/s11668-021-01205-8>
- [6] Fan, Y., Nowaczyk, S., & Rögnvaldsson, T. (2020). Transfer learning for remaining useful life prediction based on consensus self-organizing models. *Reliability Engineering & System Safety*, 203, 107098. <https://doi.org/10.1016/j.res.2020.107098>
- [7] Ferreira, C., & Gonçalves, G. (2022). Remaining Useful Life prediction and challenges: A literature review on the use of Machine Learning Methods. *Journal of Manufacturing Systems*, 63, 550–562. <https://doi.org/https://doi.org/10.1016/j.jmsy.2022.05.010>
- [8] Guo, L., Li, N., Jia, F., Lei, Y., & Lin, J. (2017). A recurrent neural network-based health indicator for remaining useful life prediction of bearings. *Neurocomputing*, 240, 98–109. <https://doi.org/https://doi.org/10.1016/j.neucom.2017.02.045>
- [9] Kang, Z., Catal, C., & Tekinerdogan, B. (2021). Remaining Useful Life (RUL) Prediction of Equipment in Production Lines Using Artificial Neural Networks. *Sensors*, 21(3), 932. <https://doi.org/10.3390/s21030932>

- [10] Karim, R., Hasan, M., Kundu, A. K., & Ave, A. A. (2023). LP SVM with A Novel Similarity Function Outperforms Powerful LP-QP-Kernel-SVM Considering Efficient Classification. *Informatica*, 47(8). <https://doi.org/10.31449/inf.v47i8.4767>
- [11] Khan, S., Yairi, T., Tsutsumi, S., & Nakasuka, S. (2024). A review of physics-based learning for system health management. *Annual Reviews in Control*, 57, 100932. <https://doi.org/https://doi.org/10.1016/j.arcontrol.2024.100932>
- [12] Kumar, H. S., & Upadhyaya, G. (2023). Fault diagnosis of rolling element bearing using continuous wavelet transform and K- nearest neighbour. *Materials Today: Proceedings*, 92, 56–60. <https://doi.org/https://doi.org/10.1016/j.matpr.2023.03.618>
- [13] Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, 104, 799–834. <https://doi.org/https://doi.org/10.1016/j.ymssp.2017.11.016>
- [14] Li, X., Elasha, F., Shanbr, S., & Mba, D. (2019). Remaining Useful Life Prediction of Rolling Element Bearings Using Supervised Machine Learning. *Energies*, 12(14), 2705. <https://doi.org/10.3390/en12142705>
- [15] Liu, A. (2024). Multi-genre Digital Music Based on Artificial Intelligence Automation Assisted Composition System. *Informatica*, 48(5). <https://doi.org/10.31449/inf.v48i5.5474>
- [16] Motahari-Nezhad, M., & Jafari, S. M. (2023). Comparison of MLP and RBF neural networks for bearing remaining useful life prediction based on acoustic emission. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, 237(1), 129–148. <https://doi.org/10.1177/13506501221106556>
- [17] Nabi, F. G., Sundaraj, K., Vijejan, V., Shafiq, M., Planiappan, R., Talib, I., & Rehman, H. U. (2021). A Novel Design of Robotic hand Based on Bird Claw Model. *Journal of Physics: Conference Series*, 1997(1), 012034. <https://doi.org/10.1088/1742-6596/1997/1/012034>
- [18] Nejjar, I., Geissmann, F., Zhao, M., Taal, C., & Fink, O. (2024). Domain adaptation via alignment of operation profile for Remaining Useful Lifetime prediction. *Reliability Engineering & System Safety*, 242, 109718. <https://doi.org/https://doi.org/10.1016/j.res.2023.109718>
- [19] Palaniappan, R., Nataraj, S. K., Ismail, Z., & Noaman, N. M. (2023a). Averaged EMG Profile and Artificial Neural Network Based Walking Speed Estimation by GAIT Analysis. *2023 IEEE 8th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, 1–8. <https://doi.org/10.1109/ICETAS59148.2023.10346388>
- [20] Palaniappan, R., Nataraj, S. K., Ismail, Z., & Noaman, N. M. (2023b). Fuzzy Logic Based Obstacle Avoidance Algorithm for Unmanned Systems. *2023 IEEE 8th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, 1–6. <https://doi.org/10.1109/ICETAS59148.2023.10346317>
- [21] Palaniappan, R., Nataraj, S. K., Noaman, N. M., & Ismail, Z. (2023). GRAFCET Based Modelling of Processing Operation in Modular Production System. *2023 IEEE 8th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, 1–5. <https://doi.org/10.1109/ICETAS59148.2023.10346373>
- [22] Rathore, M. S., & Harsha, S. P. (2022). Prognostic Analysis of High-Speed Cylindrical Roller Bearing Using Weibull Distribution and k-Nearest Neighbor. *Journal of Nondestructive Evaluation, Diagnostics and Prognostics of Engineering Systems*, 5(1). <https://doi.org/10.1115/1.4051314>
- [23] Sharma, A. K., Punj, P., Kumar, N., Das, A. K., & Kumar, A. (2024). Lifetime Prediction of a Hydraulic Pump Using ARIMA Model. *Arabian Journal for Science and Engineering*, 49(2), 1713–1725. <https://doi.org/10.1007/s13369-023-07976-6>
- [24] Yan, M., Wang, X., Wang, B., Chang, M., & Muhammad, I. (2020). Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model. *ISA Transactions*, 98, 471–482. <https://doi.org/10.1016/j.isatra.2019.08.058>
- [25] Yang, B., Liu, R., & Zio, E. (2019). Remaining Useful Life Prediction Based on a Double-Convolutional Neural Network Architecture. *IEEE Transactions on Industrial Electronics*, 66(12), 9521–9530. <https://doi.org/10.1109/TIE.2019.2924605>
- [26] Yaseen, A. S., Marhoon, A. F., & Saleem, S. A. (2022). Multimodal Machine Learning for Major League Baseball Playoff Prediction. *Informatica*, 46(6). <https://doi.org/10.31449/inf.v46i6.3864>
- [27] Zhang, L., Mu, Z., & Sun, C. (2018). Remaining Useful Life Prediction for Lithium-Ion Batteries Based on Exponential Model and Particle Filter. *IEEE Access*, 6, 17729–17740. <https://doi.org/10.1109/ACCESS.2018.2816684>
- [28] Zhang, Y., & Zhao, M. (2023). Cloud-based in-situ battery life prediction and classification using machine learning. *Energy Storage Materials*, 57, 346–359. <https://doi.org/https://doi.org/10.1016/j.ensm.2023.02.035>

- [29] Zhao, Z., Bin Liang, Wang, X., & Lu, W. (2017). Remaining useful life prediction of aircraft engine based on degradation pattern learning. *Reliability Engineering & System Safety*, 164, 74–83.
<https://doi.org/https://doi.org/10.1016/j.ress.2017.02.007>
- [30] Zhou, Y., Huang, M., & Pecht, M. (2020). Remaining useful life estimation of lithium-ion cells based on k-nearest neighbor regression with differential evolution optimization. *Journal of Cleaner Production*, 249, 119409.
<https://doi.org/https://doi.org/10.1016/j.jclepro.2019.119409>