

A unique statistical framework to predict the health of a machine by utilizing the vibration features of rolling element bearing data

Saima Bhatti ^{a*}, Fozia Shaikh ^a, Asif Mansoor ^b, Asif Ali Shaikh ^a

^a Department of Basic Science and Related Studies, Mehran University of Engineering & Technology, Jamshoro 76062, Pakistan

^b Department of Industrial & Manufacturing Engineering, PN Engineering College, National University of Sciences & Technology Karachi 07548, Pakistan

* Corresponding Author: Saima Bhatti, Email: saima.bhatti@faculty.muet.edu.pk

Received: 19 February 2025, Revised: 22 June 2025, Accepted: 30 June 2025, Published: 01 July 2025

KEYWORDS

Roller Element Bearing (REB)
Healthy-Bearing (HB)
Inner-Race-Faulty-Bearing (IRFB)
Outer-Race-Faulty-Bearing (ORFB)
Root Mean Square (RMS)
Condition Monitoring (CM)

ABSTRACT

Rolling element bearings are critical components in machinery that rotates, supporting radial and axial loads while facilitating smooth motion. Bearing malfunctions rank among the most frequent mechanical failures in rotating machinery, resulting in significant downtime and repair costs if not identified promptly. The Root Mean Square (RMS) value is a widely used statistical feature in Condition Monitoring (CM), providing a reliable and quantitative technique for detecting early-stage bearing faults. This study introduces a method for predicting machine health by utilizing RMS values derived from baseline vibration data of Healthy-Bearing (HB), Inner-Race-Faulty-Bearing (IRFB), and Outer-Race-Faulty-Bearing (ORFB) to explore the deterioration of roller bearings under various conditions using a statistical approach. The Autoregressive Moving Average (ARMA) (p,q) model was employed to fit the extracted vibration features from bearing data to create a stochastic relationship between the healthy data and the inner race faulty data, as well as between the healthy data and the outer race faulty data. The model's accuracy of the predictive model is evaluated using an error matrix. This study adds to advancing data-driven prognostics in industrial machinery by merging statistical modeling with vibration-based condition monitoring.

1. Introduction

In contemporary industrial machinery, the Rolling Element Bearing (REB) is a crucial component that significantly influences the health and longevity of machinery [1]. For rotating mechanical equipment, using fault prediction to effectively set up scheduled repairs is important to reduce maintenance costs and prevent safety incidents [2]. Rolling element bearings are essential components in many rotating machineries, directly impacting their reliability and performance. The timely detection and recognition of

REB faults are essential for enhancing the reliability of mechanical parts, preventing unexpected machine breakdowns, and lowering maintenance costs. [3-4].

In everyday situations in the industries, faulty bearings cause most of the failures in rotating equipment [5]. When the REBs develop faults, whether in the outer race, inner race, or the rolling elements themselves—they produce unique vibration signals that can be analyzed to determine the type and severity of the fault. Vibration analysis is especially useful for this, as it captures how the bearing behaves

dynamically under various fault conditions, making it an effective method for fault detection and condition monitoring [6-7]. Time-domain features include basic statistical measures such as root mean square, crest factor, peak-to-peak value, skewness, and kurtosis, which give a rapid understanding of the fault's severity and nature. Statistical techniques such as kurtosis are usually employed to detect the impulsive vibration in a bearing is due to a surface defect [8]. In vibration analysis, the Root Mean Square (RMS) value is a significant statistical feature applied to quantify the energy of vibration signals [9]. RMS offers an estimation of the overall magnitude of a signal, providing insight into the harshness of vibrations. It continues to find extensive use in condition monitoring and fault detection, particularly for rotating machines such as motors, gearboxes, and rolling element bearings. Calculating the RMS value simplifies the interpretation of complex vibration signals, which regularly contain various frequencies and magnitudes resulting from different styles of faults or operational conditions.

The RMS value is a fundamental and normally used measurement for vibration analysis; it can identify some common defects in bearings, such as inner race and outer race faults [10]. The RMS value is found to be higher than normal most of the time in the case of an inner race fault. For this reason, a defect in the inner race generates large vibrations that directly affect the bearing housing. The RMS value increases even more when the severity of the fault worsens, due to a higher impact force. In comparison, faults in the outer race create a more rhythmic pattern, often at a frequency linked to the defect on the outer race. Typically, an outer race fault will result in a higher RMS value; however, its vibration signature is generally more stable and exhibits lower amplitude than that of an inner race fault. This is because the outer race does not rotate, leading to variations in impact loads based on the load application point. Nevertheless, a continuous increase in the RMS value could suggest an outer race fault, assuming the defect is located within the load zone.

The degradation method's features are classified according to severity, reflecting the different patterns observed in financial institutions. The decline trend of the bearing element illustrated in Fig. 1 demonstrates several unique phases. FEMTO and IMS set of bearing data [11]. The rolling bearing does not fail in the health or trouble-free phase, and the RMS value varies arbitrarily. During the faulty phase or unhealthy condition, the RMS value increases as the bearing worsens.

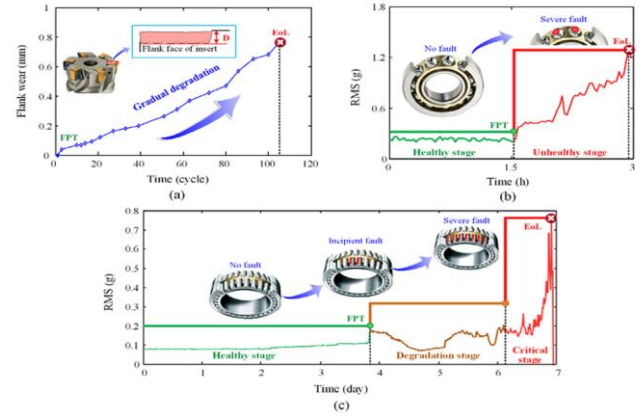


Fig. 1. FEMTO and IMS set of bearing data

2. Proposed Framework

To examine the statistical characteristics of the baseline data gathered from the experimental arrangement under different operating conditions [12]. The standard formula was utilized to derive the Root Mean Square values from healthy bearings, Inner-Race-Faulty-Bearing (IRFB), and Outer-Race-Faulty-Bearing (ORFB).

$$\text{Root Mean Square} = X = \left[\frac{\sum_{i=1}^n (x_i(t))^2}{n} \right]^{1/2} \quad (1)$$

AutoRegressive Moving Average (ARMA) modeling techniques create a connection between the identified features and normal bearings by utilizing two simulation processes: Auto-Regressive (AR) and Moving-Average (MA) models.

The AR model of order p is expressed as:

$$X_t = \alpha + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (2)$$

In order to determine future values based on lagging computations, the auto-regressive model combines lagged time-series data with error terms.

$$X_t = \alpha + \sum_{i=1}^p \phi_i L^i X_{t-i} + \varepsilon_t \quad (3)$$

The Moving-Average model is a straightforward process that selects the simplest period of the time series model. if ε_{t-j} ($j=1,2,3, \dots, q$) is a white-noise progression with zero mean and constant variance, then

$$X_t = \alpha + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \dots + \theta_q \varepsilon_{t-q} \quad (4)$$

The above equation represents the q^{th} -order MA model.

$$X_t = \alpha + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (5)$$

The lag operator might be used to rewrite the above equation as follows

$$LX_t = X_{t-1} \quad (6)$$

The moving average model can then be written to lag-operator as

$$X_t = \mu + \sum_{j=1}^q \theta_j L^j \varepsilon_{t-1} \quad (7)$$

The ARMA model of order p and q is developed by utilizing equations (3) and (7) as follows

$$X_t = \alpha + \sum_{i=1}^p \phi_i L^i X_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j L^j \varepsilon_{t-j} \quad (8)$$

Where α denotes the feature's erratic intercept terms. The ϕ_p and θ_q denote the AR and MA factors.

Finally, an error matrix or accuracy matrix concerning the actual value and statistical model forecast is applied to check the accuracy of the prediction model.

3. Methodology

In this study, we systematically analyzed various attributes across a variety of conditions and loads to elucidate the degradation of REB using statistical techniques. The extracted features of RMS are presented in Fig. 2, showing the experimental conditions under scrutiny.

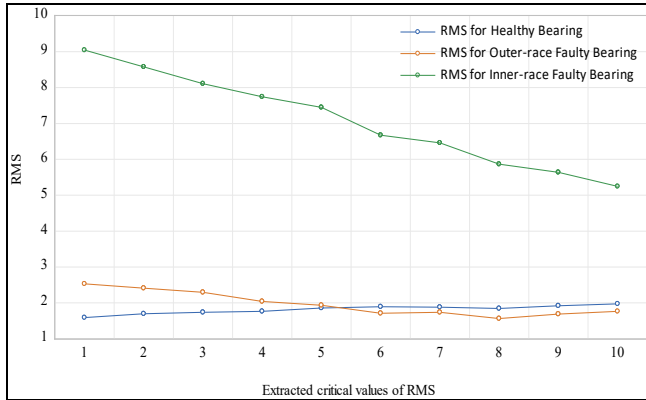


Fig. 2. Extracted RMS values

The RMS value of the vibration signal from the inner race reflects the maximum intensity of vibration, as illustrated in the graph. Conversely, the vibration metrics of the outer race and the normal condition are tightly linked, indicating a higher likelihood of machine failure when it pertains to the inner race. The features that have been extracted are utilized to build models through stochastic techniques.

To begin with, models were developed by regressing faulty features from the inner race against their related healthy features. Furthermore, the required vibration intensity factors for defects in the

outer race are treated as regressands, while corresponding healthy factors act as repressors.

Table 1

Extracted feature RMS values

S-No	HB	IRFB	ORFB
1	1.589282	9.044265	2.532688
2	1.699252	8.575836	2.412728
3	1.739354	8.109629	2.297144
4	1.765443	7.740905	2.038818
5	1.856924	7.449317	1.933012
6	1.891796	6.67444	1.710238
7	1.880441	6.458844	1.73942
8	1.844404	5.86571	1.56214
9	1.922277	5.639346	1.690271
10	1.972779	5.244006	1.763291

Equation (8), along with extracted RMS values statistics shown in Table 1, are used to develop statistical models of RMS for inner race fault and outer race fault, respectively shown below:

a. Statistical Model of RMS for Inner Race Fault

$$\left. \begin{aligned} \Phi_t^i &= 17.96 - 6.037\Phi_t^h + 0.84\Phi_{t-1}^h + \omega_t^{ri} \\ p\text{-values: } &(0.012) \quad (0.091) \quad (0.076) \\ R^2 &= 0.91 \end{aligned} \right\} \quad (9)$$

b. Stochastic Model of RMS for Outer Race Fault

$$\left. \begin{aligned} \Phi_t^o &= 7.38 - 2.987\Phi_t^h - 0.524(\Phi_{t-1}^h + \Phi_{t-2}^h) + \omega_t^{ro} \\ p\text{-values: } &(0.001) \quad (0.002) \quad (0.089) \\ R^2 &= 0.84 \end{aligned} \right\} \quad (10)$$

Where $\Phi_t^i, \Phi_t^o, \Phi_t^h$ represent the RMS for IRFB, ORFB, and Healthy state, ω_t^{ri} and ω_t^{ro} represent the inner race and outer-race RMS white-noise-term with corresponding variance.

The p-values of the coefficients in equations (9) and (10) indicate good significance of the models. The values of R^2 for inner and outer race faults are 84%, and 91%, indicate the variance between the model value at time t and the independent variables. Overall, the R^2 values show that the models are statistically significant.

ARMA modeling methods are utilized to identify optimized parameters to develop a strong connection between malfunctioning signals and healthy bearings. Table 2 represents the optimized ordering of the statistical models

The RMS value, a commonly utilized statistical indicator when a fault is present, the complexity of the vibration signal, leading to alterations in the RMS value and variations in the estimated coefficients.

Table 2

Order of the Suitable model

Models	Order	AIC	BIC
RMS of IRFB	ARMA (1,0)	1.7768	1.8978
RMS of ORFB	ARMA (2,0)	-0.4051	-0.2341

4. Results and Discussion

To evaluate the performance of the statistical models the error matrix is commonly used. The accuracy and reliability of the models developed to identify traits linked to inner and outer race faults were evaluated by employing an error matrix that contrasts the observed extracted features with the values produced by the model. Table 3 presents the crucial forecast statistics of the recognized RMS model of IRFBs, indicating its consistency. Overall, the statistical results indicate that the predicted model provides quite accurate forecasts.

Table 3

Result of the accuracy matrix of the RMS Model of IRFB

Forecast Statistics	Estimating Parameter	Acceptance Range Value
RMSE	0.3716	5% to 10%
MAE	0.2980	5% to 10%
MAPE	4.572	< 10%
Theil Coefficient	0.0265	0 to 1
SMAPE	4.442	< 10%

The prediction result obtained from the RMS model inner race faulty bearing with a standard confidence interval is shown in Fig. 3.

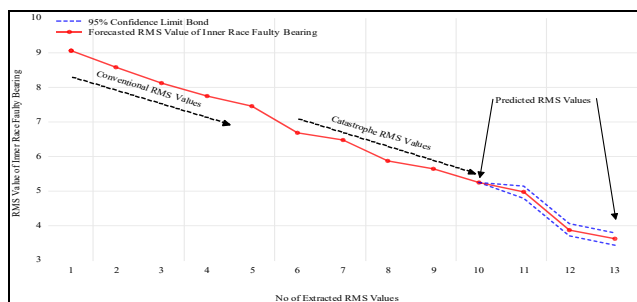


Fig. 3. Forecast Results for RMS of the IRFB with 95% CI.

Under standard conditions, there is minimal amplitude, indicating a healthy system with minimal vibrations. It demonstrates smooth movement with few issues. The defect in the inner race leads to a gradual rise in the RMS value. This slight variation stems from the introduction of imprecise movement in the rolling elements, which enhances the transfer of

vibrational energy throughout the machine structure. The RMS value continues to rise as the inner race defect worsens. Increased vibrational energy indicates more severe fault conditions that could cause greater impacts on the rolling elements and damage to the inner race surface.

Table 4 shows a comprehensive view of the RMS model's performance for outer race faulty features. The accuracy matrix comparing the extracted observed features and the values produced by the developed model determines the accuracy of the model. All statistical forecast values fall within an acceptable range, indicating a high level of accuracy for the model. Furthermore, continuously monitoring and evaluating these statistics is crucial for successful predictive maintenance in industrial situations.

Table 4

Result of the accuracy matrix of the RMS Model of ORFB

Forecast Statistics	Estimating Parameter	Acceptance Range Value
RMSE	0.1363	5% to 10%
MAE	0.0975	5% to 10%
MAPE	5.6463	< 10%
Theil Coefficient	0.036	0 to 1
SMAPE	5.4769	< 10%

The outcomes predicted by the RMS model regarding the features of an outer race fault are illustrated in Fig 4, with the RMS model projecting within a 95% confidence interval for both lower and upper bounds.

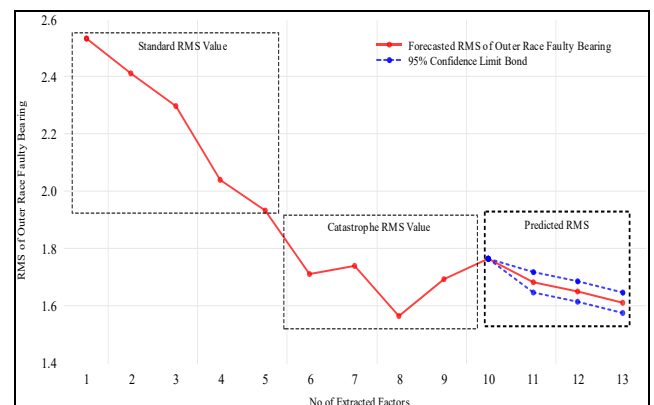


Fig. 4. Forecast Results for RMS of the ORFB with 95% CI.

The RMS model illustrates the connection between the RMS value and faults in the outer race, providing valuable insights into the development of bearing problems. The rise in vibration energy within the RMS model suggests the possible dynamics of failure.

5. Conclusion

The statistical and forecasting outcomes discussed in this research are used to evaluate the performance of the machinery model across different experimental conditions. The results of RMS models for IRFB and ORFB were analyzed with an error matrix bounded to the 95% confidence interval (CI) to show the prediction precision. Overall, the forecast tables and prediction graphs validate the model's consistency, as indicated by the forecast accuracy values in the associated tables. This forms a unified connection between visual depiction and quantitative evaluation. Essentially, both the visual and tabular analyses come together to reinforce the finding that the developed model demonstrates reliability and precision in forecasting the Remaining Useful Life (RUL) of machinery under various experimental situations.

6. References

- [1] I. El-Thalji and E. Jantunen, "Fault analysis of the wear fault development in rolling bearings," *Engineering Failure Analysis*, vol. 57, pp. 470–482, 2015.
- [2] H. Li, Y. Zhang, S. Wang, K. Gu, and M. Ren, "Multi-sensor vibration signal-based three-stage fault prediction for rotating mechanical equipment," *Entropy*, vol. 24, no. 2, p. 164, 2022.
- [3] S. A. Aburakhia, "Data-Driven Vibration-Based Condition Monitoring: Fundamentals, Applications, and Challenges", Doctoral dissertation, The University of Western Ontario (Canada), 2024.
- [4] P. Jha, S. Mahato, P. K. Jana, S. Maurya, and I. Chihi, "Artificial Intelligence-based Solutions for Industrial Applications". CRC Press, 2024.
- [5] M. Hussain, A. Mansoor, and S. Nisar, "Bearing Degradation Prognosis Using Structural Brak Classifier," *Mechanika*, vol. 24, no. 4, pp. 456–461, 2018. [Online]. Available: <https://doi.org/10.5755/j01.mech.4.24.20740>
- [6] M. Hussain, A. Mansoor, and Q. Ali Rehman, "A Novel Approach for Machinery Health Prognostics Using Statistical Tools," in *Proc. 50th MFPT Conf.*, 2017, pp. 1–13.
- [7] X. Liu, L. Bo, X. He, and M. Veidt, "Application of correlation matching for automatic bearing fault diagnosis," *Journal of Sound and Vibration*, vol. 331, pp. 5838–5852, 2012.
- [8] W. Li, M. Qiu, Z. Zhu, B. Wu, and G. Zhou, "Bearing fault diagnosis based on spectrum images of vibration signals," *Measurement Science and Technology*, vol. 27, p. 035005, 2016.
- [9] S. Bhatti, A. A. Shaikh, A. Mansoor, and M. Hussain, "New Approaches of Stochastic Models to Examine the Vibration Features in Roller Bearings," **Applied Sciences**, vol. 14, p. 1616, 2024. [Online]. Available: <https://doi.org/10.3390/app14041616>
- [10] C. Bianchini, F. Immovilli, M. Cocconcelli, R. Rubini, and A. Bellini, "Fault detection of linear bearings in brushless AC linear motors by vibration analysis," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 1684–1694, 2010.
- [11] T. O. Hodson, "Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not," *Geosci. Model Dev. Discuss.*, pp. 1–10, 2022.
- [12] S. Bhatti, A. A. Shaikh, A. Mansoor, and R. Shaikh, "Extraction and classification of vibration features of rolling element bearing with increasing rotational speed," *Mehran Univ. Res. J. Eng. Technol.*, vol. 43, no. 4, pp. 182–191, 2024.