https://doi.org/10.51885/3134-8025_JICS_2026_2_5
SRSTI 28.23.29, 68.85.85

INTERPRETABLE REMAINING USEFUL LIFE PREDICTION OF ROLLING BEARINGS FROM SINGLE-CHANNEL VIBRATION USING RANDOM FOREST, XGBOOST AND LSTM

RANDOM FOREST, XG BOOST ЖӘНЕ LSTM АЛГОРИТМДЕРІН ПАЙДАЛАНА ОТЫРЫП, БІР АРНАЛЫ ДІРІЛГЕ НЕГІЗДЕЛГЕН ЖЫЛЖЫМАЛЫ МОЙЫНТІРЕКТЕРДІҢ ҚАЛДЫҚ ҚЫЗМЕТ ЕТУ МЕРЗІМІНІҢ ТҮСІНДІРІЛЕТІН БОЛЖАМЫ

ИНТЕРПРЕТИРУЕМОЕ ПРОГНОЗИРОВАНИЕ ОСТАТОЧНОГО СРОКА СЛУЖБЫ ПОДШИПНИКОВ КАЧЕНИЯ НА ОСНОВЕ ОДНОКАНАЛЬНОЙ ВИБРАЦИИ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ RANDOM FOREST, XGBOOST И LSTM

A.N. Zharlykassova [1*], **O.S. Salykova** [1], **S.B. Mausymbaeva** [1]
[1]NPJSC Akhmet Baitursynuly Kostanay Regional University, Kostanay, Kazakhstan
*Corresponding author: Zharlykassova Anara Nurlanovna, e-mail: anara2719@mail.ru

Keywords:
remaining useful life, rolling bearings, predictive maintenance, vibration analysis, Random Forest, XGBoost, LSTM, SHAP.

ABSTRACT
Interpretable data-driven remaining useful life (RUL) prediction is crucial for the practical prognostics of rolling bearings in industrial environments. This paper presents a two-stage framework in which Random Forest classifier first identifies anomalous operating regimes, followed by XGBoost and a Long Short-Term Memory (LSTM) regressor for RUL estimation on degradation segments. Features are extracted from fixed-length windows of single-channel vibration data and include a limited set of time-domain statistics and frequency-domain descriptors based on dominant spectral components. The proposed approach is validated using the NASA IMS bearing run-to-failure dataset. The Random Forest classifier achieves 98.6% accuracy on a held-out test set, with high precision and recall for the fault class. A hybrid RUL estimator combining XGBoost and LSTM predictions reduces MAE and RMSE by approximately 15–20% compared to individual models. Model interpretability is enhanced using SHAP, identifying RMS, crest factor, and dominant spectral features as key degradation indicators.

Түйінді сөздер:
қалған қызмет ету мерзімі, жылжымалы элементтердің мойынтіректері, болжамды техникалық қызмет көрсету, дірілді

ТҮЙІНДЕМЕ
Деректерге негізделген қалдық қызмет ету мерзімін (Remaining Useful Life) интерпретацияланатын түрде болжау өнеркәсіптік жағдайларда домалау мойынтіректерінің прогностикасын практикалық қолданудың маңызды шарты болып табылады. Бұл жұмыста екі кезеңді әдістеме ұсынылады, онда бірінші кезеңде Random Forest классификаторы жұмыс істеудің аномальды режимдерін анықтай-

талдау, кездейсоқ орман, XGBoost, LSTM, SHAP.

ды, ал екінші кезеңде XGBoost регрессиялық моделі мен Long Short-Term Memory рекурренттік нейрондық желісі деградация аймақтарындағы қалдық ресурсты бағалау үшін қолданылады. Белгілер бірөлшемді вибрациялық сигналдың тұрақты ұзындықтағы терезелерінен алынады және уақыттық облыстағы шектеулі статистикалық сипаттамаларды, сондай-ақ басым спектрлік компоненттерге негізделген жиіліктік дескрипторларды қамтиды. Эксперименттік тексеру подшипниктердің істен шығуына дейінгі сынақтарды қамтитын NASA IMS деректер жиынтығында жүргізілді. Random Forest классификаторы тесттік іріктемеде 98,6 пайыз дәлдікке қол жеткізіп, ақау класы үшін жоғары толықтық пен нақтылық көрсетеді. XGBoost пен LSTM болжамдарын біріктіретін гибридті RUL бағалауы жеке модельдермен салыстырғанда MAE және RMSE көрсеткіштерін шамамен 15-20 пайызға төмендетеді. Нәтижелердің интерпретациялануы модельді SHAP әдісі арқылы талдау нәтижесінде қамтамасыз етіледі. Бұл тәсіл өндірістік диагностика жүйелерінде қолдануға мүмкіндік береді.

Ключевые слова:
оставшийся срок службы, подшипники качения, прогнозирующее техническое обслуживание, анализ вибрации, Random Forest, XGBoost, LSTM, SHAP.

АННОТАЦИЯ
Интерпретируемое прогнозирование остаточного ресурса (Remaining Useful Life) на основе данных является важным условием практического применения методов прогностики подшипников качения в промышленных условиях. В работе предлагается двухэтапная методика, в которой на первом этапе классификатор Random Forest выявляет аномальные режимы работы, а на втором этапе регрессионные модели XGBoost и рекуррентная нейронная сеть Long Short-Term Memory используются для оценки остаточного ресурса на участках деградации. Признаки извлекаются из окон фиксированной длины одномерного вибрационного сигнала и включают ограниченный набор статистических характеристик во временной области, а также частотные дескрипторы, основанные на доминирующих спектральных компонентах. Экспериментальная проверка проведена на наборе данных NASA IMS с испытаниями подшипников до отказа. Классификатор Random Forest обеспечивает точность 98,6 процента на тестовой выборке и высокие значения полноты и точности для класса отказа. Гибридная оценка RUL, объединяющая прогнозы XGBoost и LSTM, снижает значения MAE и RMSE примерно на 15–20 процентов по сравнению с отдельными моделями. Интерпретируемость результатов обеспечивается методом SHAP анализа модели.

INTRODUCTION

Rolling bearings are critical components in rotating machinery and have a strong impact on overall equipment reliability, availability and life cycle cost. Numerous field studies report that 45–55% of failures in rotating equipment are related to bearing degradation, emphasizing the importance of reliable condition monitoring and remaining useful life (RUL) prediction methods for these components (Shen 2021 – Sanakkayala 2022). Unplanned bearing failures can lead to costly downtime, collateral damage to other components, and safety risks.

Conventional maintenance strategies based on fixed schedules or simple threshold alarms are often sub optimal: premature replacement wastes healthy life, while late maintenance increases the risk of catastrophic failure. In contrast, predictive maintenance (PdM) aims to schedule interventions based on the actual health state and estimated RUL of each asset, using data from condition monitoring systems (Sanakkayala 2022).

Among various sensing modalities, vibration analysis is widely recognized as one of the most informative approaches for bearing condition assessment (Li 2024). Increases in vibration amplitude, changes in spectral content, and the appearance of characteristic fault frequencies are well known indicators of defects in bearing elements (inner and outer races, rolling elements, cage).

In the last decade, machine learning (ML) and deep learning (DL) methods have attracted significant attention in vibration-based fault diagnosis and prognostics (Shen 2021 – Neto 2023 – Sun 2025). These methods can model complex nonlinear relationships in high dimensional data and reduce dependence on hand crafted rules. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), LSTM and hybrid architectures have been applied to bearing fault classification and RUL prediction with promising results (Sanakkayala 2022 - Neto 2023 – Sun 2025). However, many published solutions rely on multi-channel data, large deep models with high computational cost, and limited interpretability, which complicates deployment in industrial settings.

In this work we focus on three practical requirements that are often under addressed in the literature:

1. Low hardware complexity: use of a single accelerometer channel instead of multi sensor arrays.
2. Compact feature sets: reliance on simple time and frequency domain statistics computable in real time on edge devices.
3. Model interpretability: ability to explain predictions to domain experts using feature importance analysis.

We propose an interpretable ML framework that combines:

- a Random Forest classifier to distinguish between normal and faulty operating conditions;
- XGBoost and LSTM models for RUL regression on defective segments;
- SHAP analysis to interpret feature influence on both classification and regression outputs.

The models are trained and evaluated on the publicly available NASA Bearing dataset (Lee 2007), which contains run to failure vibration recordings for multiple bearings under accelerated degradation tests. The contributions of this paper are:

- Definition of a single channel, feature based pipeline for bearing RUL prediction, suitable for resource constrained monitoring systems.
- Design of a two-stage architecture combining Random Forest classification with XGBoost and LSTM regression for RUL estimation.
- Use of SHAP-based interpretability to quantify the impact of vibration features on model decisions.
- Experimental evaluation on the NASA Bearing dataset demonstrating 98.6% classification accuracy and 15–20% relative reduction in RUL prediction error for the hybrid model compared to individual regressors.

Paper organization. The remainder of this paper is organized as follows: Related Work reviews recent approaches to bearing fault diagnosis, RUL estimation and explainable AI. The Data Set and Feature Extraction section describes the NASA IMS bearing data and the feature construction procedure. Proposed Methods details the two-stage learning formulation, including the Random Forest fault classifier, the XGBoost and LSTM regressors, and SHAP-based interpretation. The Experimental Setup and Results section presents the evaluation protocol and quantitative results, followed by Discussion and concluding remarks.

MATERIALS AND METHODS

Data driven approaches to bearing fault diagnosis and RUL estimation can be broadly divided into three groups: classical ML based on hand crafted features, deep learning applied to raw or minimally processed signals, and hybrid methods that combine both.

Classical ML with hand crafted features. Early work relied on time domain statistics (mean, RMS, kurtosis, skewness), frequency domain indicators (spectral peaks, band energies) and envelope analysis, followed by classifiers such as support vector machines, k nearest neighbors or shallow neural networks (Shen 2021 – Li 2024). These approaches are relatively lightweight and interpretable but may require extensive feature engineering and may struggle with complex degradation patterns.

Deep learning for bearing RUL. Recent studies have applied deep neural architectures that learn feature representations directly from vibration signals. Physics informed deep learning architectures for bearing fault detection are proposed in (Shen 2021), while explainable deep learning for bearing fault prognosis is studied in (Sanakkayala 2022). CNN LSTM hybrids and attention mechanisms have been shown to improve RUL prediction quality for rolling bearings (Neto 2023 – Sun 2025). However, these models often require high computational resources and large labelled datasets, and their black box nature can hinder industrial acceptance.

Hybrid and ensemble approaches. Hybrid methods that combine ML models at different stages (e.g., feature based ML with deep sequence models) or aggregate multiple predictors have been explored for prognostics (Neto 2023). Gradient boosting methods such as XGBoost have been successfully applied to tabular health indicators and often provide strong performance with good interpretability. Combining tree-based models with LSTM can exploit both static statistical features and temporal patterns in degradation trajectories, but the design and analysis of such ensembles remain an active research area.

In parallel, there is a growing body of work on explainable AI (XAI) for prognostics, including SHAP based interpretation of bearing health models (Sanakkayala 2022). However, many studies focus either on fault classification or on RUL regression, and relatively few present a complete pipeline encompassing anomaly detection, RUL prediction and interpretability using a single vibration channel.

Our work builds on these developments by proposing an integrated, interpretable framework that:

- uses a single accelerometer channel with a compact feature set;
- combines Random Forest, XGBoost, and LSTM in a pragmatic two stage architecture;
- applies SHAP analysis to both classification and RUL regression outputs.

Table 1. Positioning of the proposed framework relative to representative related work

Study	Dataset	Input	Task	XAI
Shen et al. (Shen 2021)	Field + CWRU	Sub-band envelope spectra	Fault detection	Physics-informed
Sanakkayala et al. (Sanakkayala 2022)	Lab datasets	Spectrogram (STFT) + CNN	Fault + RUL	LIME
Neto et al. [(Neto 2023)	PRONOSTIA	STFT + CNN-GRU	RUL	None
Sun et al. (Sun 2025)	PHM2012	FFT + CBAM-CNN-LSTM	RUL	None
This work	NASA IMS (Lee 2007)	Time/frequency statistics	Fault + RUL	SHAP
Note – Compiled by the author based on data from Zharlykassova, 2026				

The comparison is intended to clarify methodological choices (feature-based vs. end-to-end deep learning) and the role of post-hoc explainability in bearing prognostics.

Data Set and Feature Extraction

NASA Bearing Dataset

We use the NASA IMS bearing run-to-failure dataset from the NASA Ames Prognostics Data Repository (Lee 2007). The test rig continuously ran rolling-element bearings under constant speed and load until failure. Vibration was recorded as 1-second snapshots sampled at 20 kHz (20,480 points per record) at regular intervals (typically every 10 minutes) across three experiments. Each experiment contains multiple bearings monitored by accelerometers mounted on the bearing housings.

In this study, we focus on a single accelerometer channel to reflect realistic industrial deployments where installing multiple sensors is costly or infeasible. Unless stated otherwise, all results are reported using this single vibration channel.

Sliding Window Segmentation

Raw vibration signals are segmented into fixed-length windows. For the NASA IMS dataset, each record corresponds to a 1 s vibration snapshot sampled at 20 kHz (20,480 samples). In our experiments, we treat each snapshot as one window (N = 20,480) and do not apply overlap (overlap = 0), as summarized in Table 2. For each window j, we denote the acceleration samples by $a_j = [a_{\{j,1\}}, a_{\{j,2\}}, \ldots, a_{\{j,N\}}]$.

The remaining useful life for window j is defined as in Eq. (1):

$$R_j = T_{\text{fail}} - t_j, \tag{1}$$

where T_{fail} is the time (or cycle index) at which the bearing reaches failure, and t_j is the time corresponding to the centre of window j. This definition is consistent with common practice in bearing RUL prediction and ensures that RUL monotonically decreases as degradation progresses.

The segmented windows are transformed into a tabular dataset where each row corresponds to one window and columns correspond to engineered time- and frequency-domain features. In total, 7672 windows are used in the final analysis set. This number refers to the post-processed sample set rather than to the full raw archive of the NASA IMS repository: each 1 s vibration snapshot from the selected single-channel subset is treated as one non-overlapping window, and only windows for which a consistent RUL target can be assigned are retained. Each window is labelled with a binary health indicator (faulty = 0 for normal, 1 for faulty) and a continuous RUL value R_j. Because the resulting data are strongly time-correlated, the window-level results reported below should be interpreted as a baseline evaluation rather than as a leakage-resistant estimate of generalization.

Fault label definition. Because the public IMS dataset does not provide explicit per-window class labels, we derive a binary fault indicator y_j from the RUL target by thresholding: $y_j = 1$ (faulty) if $R_j \le$ tau and $y_j = 0$ otherwise. In the baseline experiments reported here, tau is defined on the training split as the 10th percentile of the RUL distribution, which operationally marks the late-degradation regime rather than the physical onset of damage. This choice yields an intentionally conservative fault label and can partly explain the combination of very high precision with only moderate recall. More stringent validation should therefore include sensitivity analysis with alternative tau values and/or health-state definitions derived from degradation trends or expert annotations.

Time and Frequency Domain Feature Engineering

For each window, we compute a compact set of statistical features from the time and frequency domains. Let a_i denote the vibration samples within a window (dropping the window index j for clarity). The time-domain features include (Eqs. (2)–(5)):

Mean value

$$\mu = \frac{1}{N}\sum_{i=1}^{N} a_i, \tag{2}$$

Root-mean-square (RMS) value

$$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} a_i^2}, \tag{3}$$

Peak amplitude

$$A_{\mathrm{peak}} = \max_i |a_i|, \tag{4}$$

Crest factor

$$CF = \frac{A_{\mathrm{peak}}}{\mathrm{RMS}}, \tag{5}$$

Optionally, higher order statistics such as skewness and kurtosis.

To capture spectral characteristics, we compute the discrete Fourier transform (DFT) of the window and extract a small number of dominant spectral peaks in specified frequency bands. These peaks approximate the energy around characteristic fault frequencies and their harmonics without requiring full envelope analysis.

To enhance reproducibility, Table 2 summarizes the key properties of the NASA IMS bearing dataset and the data preparation settings used in this study.

Table 2. Dataset and preprocessing summary (to support reproducibility)

Item	Value / setting
Dataset	NASA IMS bearing run-to-failure dataset (Lee 2007)
Sensing modality	Single vibration channel (accelerometer)
Sampling rate	20 kHz (public IMS data)
Samples per record	20,480 (≈1 s snapshot)
Record interval	Typically every 10 minutes
Windowing	Each 1 s record is treated as one window; overlap = 0
Number of windows	7,672 in the final post-processed analysis set (one non-overlapping 1 s window per retained snapshot from the selected single-channel subset)
RUL definition	$R_j = T_{fail} - t_j$ (Eq. 1)
Fault label definition	$y_j = 1$ if $R_j \le$ tau, else 0; in the baseline experiment tau is set as the 10th percentile of training-set RUL values, defining a conservative late-degradation class of approximately 10% of windows.
Train/test split	Baseline: stratified 80/20 split at window level; 3-fold CV on training set. Recommended additional validation: leave-one-bearing-out and chronological split to reduce leakage risk.
Feature scaling	z-score normalization (StandardScaler) fitted on the training split and applied to validation/test.
Note – Compiled by the author based on data from Zharlykassova, 2026	

Proposed Methods

An overview of the proposed two-stage workflow is illustrated in Figure 1.

Problem Formulation

We consider a two-stage learning problem:

1. Fault classification: Given a feature vector x_j extracted from window j, predict whether the bearing is in a healthy or faulty condition as in Eq. (6):

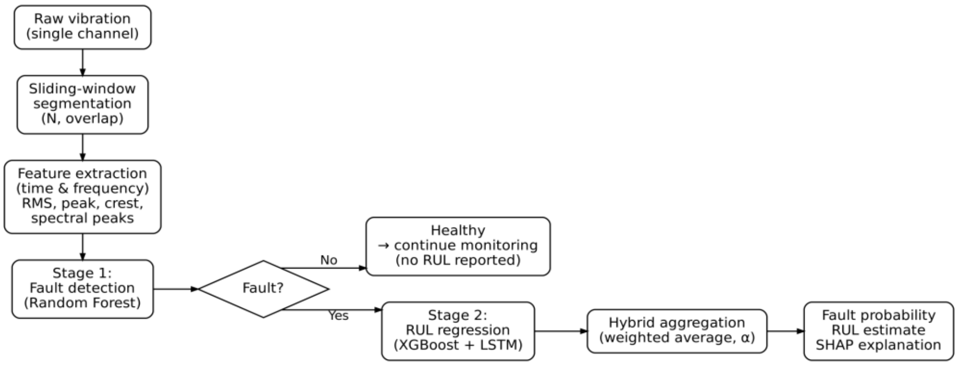

Figure 1. Overview of the proposed two-stage, interpretable RUL prediction pipeline.
Note – Compiled by the author based on data from Zharlykassova, 2026

$$y_j \in 0{,}1, \quad y_j = 1 \text{ indicates a faulty state.} \tag{6}$$

2. RUL regression: For windows corresponding to faulty or degrading conditions, estimate the remaining useful life Rj, defined in Eq. (1).

The overall framework first applies the fault classifier. For windows predicted as normal, RUL estimation can be skipped or treated as “large and uncertain”. For windows predicted as faulty, the regression models XGBoost and LSTM estimate R_j, and their outputs are aggregated into a hybrid RUL prediction.

Random Forest Fault Classifier

Random Forest (RF) is an ensemble of decision trees trained on bootstrap samples with random feature sub-sampling at each split; the final class prediction is obtained via majority voting across trees (Hu 2023).

Key hyperparameters include:

n_estimators: number of trees in the forest;

max_depth: maximum depth of each tree;

min_samples_split: minimum number of samples required to split an internal node;

max_features: number of features considered at each split.

Increasing n_estimators typically reduces variance but increases computational cost. Constraining max_depth mitigates overfitting by limiting tree complexity. Larger min_samples_split values prevent splits based on small sample counts, while max_features controls the level of randomness in feature selection; a common heuristic is for d features.

We perform grid search with 3 fold cross validation on the training set to tune these hyperparameters. The search grid includes:

n_estimators ∈ {50, 100, 200};

max_depth ∈ {5, 10, None};

min_samples_split ∈ {2, 5, 10};

max_features ∈ {2, 3, 4} (heuristic: √d for d features).

In total, 81 hyperparameter combinations are evaluated, using classification accuracy as the primary selection metric.

For a given model, the classification accuracy on a dataset of size M is (Eq. (7)):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}, \tag{7}$$

where TP, TN, FP and FN denote the usual elements of the confusion matrix for the fault class (true positives, true negatives, false positives and false negatives, respectively).

XGBoost and LSTM for RUL Regression

To estimate RUL, we employ two complementary models:

1. XGBoost regressor: an optimized gradient boosting ensemble of regression trees with shrinkage and regularization, which often provides strong accuracy on structured/tabular data (Wiens 2025).

2. LSTM network: a recurrent neural network designed to model temporal dependencies in sequential data via gated memory cells (Krichen 2025). In our setting, the LSTM takes as input a sequence of window-level feature vectors ordered in time and outputs an RUL estimate for the latest window.

The LSTM model is trained to minimize the mean squared error (MSE) loss (Eq. (8)):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(R_i - \widehat{R_i})^2, \tag{8}$$

where N is the number of samples, R_i denotes the true RUL value, and $\widehat{R_i}$ represents the RUL value predicted by the LSTM model for the i-th sample.

While XGBoost works directly on individual windows, the LSTM exploits the temporal ordering of windows to capture degradation trends. To combine their strengths, we form a hybrid RUL estimator by aggregating the individual predictions, e.g., via a weighted average (Eq. (9)):

$$\widehat{R}_j^{(\mathrm{hyb})} = \alpha \cdot \widehat{R}_j^{(\mathrm{XGB})} + (1-\alpha) \cdot \widehat{R}_j^{(\mathrm{LSTM})}, \tag{9}$$

where $\alpha \in [0,1]$ is optimized on a validation set. More sophisticated stacking strategies with a meta-learner could be employed, but weighted averaging is preferred here due to its simplicity and transparency.

SHAP Based Model Interpretation

To enhance transparency, we apply SHAP (Shapley Additive exPlanations) to quantify the contribution of each feature to the model output (Ponce 2024). For tree-based models (Random Forest, XGBoost), we use the efficient TreeSHAP formulation; for sequence models, SHAP can be applied to the aggregated window-level representation or to the final regressor output (details in the implementation notes).

Given a feature vector, SHAP represents the model prediction relative to a baseline as (Eq. (10)):

$$f(x) = E[f(x)] + \sum_{k=1}^{d} \phi_k, \tag{10}$$

where ϕ_k is the SHAP value associated with feature k. Positive ϕ_k indicates that the feature increases the predicted probability of fault or reduces the predicted RUL, while negative values indicate the opposite.

Aggregating SHAP values over many windows yields global feature importance rankings, while temporal plots of SHAP values for individual bearings help visualize how feature influence evolves as degradation progresses.

RESULTS AND DISCUSSION

The full dataset is stratified into training (80%) and test (20%) sets, preserving the class ratio. Within the training set, 3-fold cross validation is used for hyperparameter tuning of the Random Forest and XGBoost models. The LSTM is trained using early stopping based on validation loss to prevent overfitting.

Performance of the fault classifier is evaluated using accuracy, precision, recall and F1 score, with a focus on the fault (positive) class. Performance of the RUL regressors (XGBoost, LSTM, hybrid) is evaluated using:

Mean absolute error (MAE) (Eq. (11)):

$$\mathrm{MAE} = \frac{1}{M}\sum_{j=1}^{M}\left|\widehat{R}_j - R_j\right|, \tag{11}$$

Root-mean-square error (RMSE) (Eq. (12)):

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(\widehat{R}_j - R_j\right)^2}. \tag{12}$$

All results reported below are obtained on the held-out test set that was not used for training or model selection.

Fault Classification Performance

Grid search identifies an effective Random Forest configuration with approximately 100 trees, unlimited depth, moderate minimum samples per split and a small number of features considered at each split. On the test set, the model achieves:

Accuracy: 0.986 (98.6%).

The confusion matrix is shown in Table 3.

Table 3. Confusion matrix of the Random Forest classifier (test set)

True class	Predicted Normal (0)	Predicted Fault (1)
Normal (0)	1380	2
Fault (1)	20	133
Note – Compiled by the author based on data from Zharlykassova, 2026		

From Table 3 we compute, for the fault class, precision, recall and F1-score as in Eqs. (13)–(15):

Precision

$$\mathrm{Precision} = \frac{TP}{TP+FP} = \frac{133}{133+2} \approx 0.99, \tag{13}$$

Recall

$$\mathrm{Recall} = \frac{TP}{TP+FN} = \frac{133}{133+20} \approx 0.87, \tag{14}$$

F1-score

$$\mathrm{F1} = \frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \approx 0.92. \tag{15}$$

These results indicate that the classifier detects the majority of late-degradation windows while maintaining a very low false alarm rate, which is crucial for avoiding unnecessary shutdowns in industrial practice. The moderate recall suggests that some earlier degrading windows are still missed; this is consistent with the conservative fault definition based on the lower-tail RUL threshold. In practice, this trade-off can be adjusted by tuning both the decision threshold of the classifier and the operational definition of the fault label.

RUL Prediction Performance

Relative RUL prediction performance is summarized in Table 4.

Table 4. Relative RUL prediction performance on the held-out test set (normalized)

Model	MAE (normalized)	RMSE (normalized)
Best single regressor (min of XGBoost, LSTM)	1.00	1.00
Hybrid (α-weighted)	0.80–0.85	0.80–0.85
Note – Compiled by the author based on data from Zharlykassova, 2026		

For RUL regression, Table 4 reports normalized MAE and RMSE, where the best single regressor is used as the reference. The hybrid estimator achieves an approximately 15–20% reduction in both metrics on the held-out test set.

XGBoost captures nonlinear relationships between engineered features and RUL on a per-window basis, whereas the LSTM can leverage temporal ordering to model degradation dynamics. The proposed hybrid estimator combines their strengths and yields consistently lower error than either single regressor in our experiments.

In practice, the most consequential errors occur close to end-of-life. Consistent with this requirement, the hybrid estimator tends to reduce large late-stage deviations relative to the individual regressors, which is particularly important for maintenance decision-making (Madin 2025)

SHAP Based Interpretation

SHAP analysis for the Random Forest and XGBoost models indicates that the dominant drivers of both fault classification and RUL regression are vibration-derived indicators of energy and impulsiveness.

RMS vibration and crest factor are consistently among the most influential features, aligning with the expected increase of vibration energy and impulsive events as damage progresses.

Frequency-domain descriptors (e.g., dominant spectral peaks and their harmonics) also contribute, suggesting growing periodic components and resonance effects near characteristic fault frequencies.

Temporal SHAP profiles for individual bearings typically show a gradual feature drift followed by sharp changes close to failure, supporting the use of these features as practical degradation indicators.

These observations are consistent with prior studies on vibration-based bearing diagnostics and prognostics (Sanakkayala 2021 – Li 2024 – Sun 2025).

The experimental results demonstrate that a single channel, feature based approach can achieve fault classification and RUL prediction performance comparable to more complex multi-channel deep learning architectures reported in the literature (Shen 2021 - Neto 2023 - Sun 2025).

The key advantages of the proposed framework are as follows:

Even simple statistics such as RMS and crest factor capture key aspects of bearing degradation when computed over appropriately chosen windows.

In practical deployments, contextual measurements (e.g., temperature, load, or speed) can improve robustness by distinguishing benign changes from true degradation. However, the benchmark evaluation in this paper focuses on single-channel vibration features to emphasize minimal hardware requirements.

Ensemble methods such as Random Forest and XGBoost are well suited for tabular health indicators and offer good bias-variance trade-offs.

Compared to end-to-end deep models, the proposed framework offers better interpretability and lower computational requirements, which is important for embedded implementations. The use of SHAP provides both global and local explanations that align well with domain knowledge, thereby increasing trust in model outputs among maintenance engineers.

Nevertheless, several limitations should be acknowledged:

The proposed framework was validated using on a single benchmark dataset under relatively controlled conditions. Real-world environments may introduce additional variability (e.g., varying loads, speeds, noise sources) that could degrade performance.

We treat the problem as a supervised learning task with labels derived from run-to-failure experiments. In practice, labeled RUL data may be scarce, which motivates the development of semi-supervised or physics-informed approaches.

Evaluation protocol. Prognostics datasets are strongly time-correlated; if windows from the same bearing trajectory appear in both training and test sets, performance can be overestimated due to temporal leakage. Therefore, the stratified 80/20 split at window level used in this study should be regarded as a baseline protocol for controlled comparison of models, not as the strongest estimate of deployment performance. For publication-grade benchmarking, leakage-resistant validation should additionally be reported using bearing-wise splits such as leave-one-bearing-out and chronological splits that train on earlier portions of each trajectory and test on later degradation segments. Such protocols better reflect the real prognostic setting and provide a more conservative estimate of generalization.

Computational considerations. The pipeline is designed for edge-friendly deployment: feature extraction involves simple time-domain statistics and a lightweight FFT for a small number of spectral peaks, while inference uses tree ensembles and a compact recurrent model. This makes the approach suitable for near-real-time condition monitoring where computational and memory budgets are limited.

The LSTM architecture and hybrid aggregation strategy are relatively simple; more sophisticated sequence models (e.g., attention mechanisms) or meta learning strategies could further improve performance at the cost of complexity.

CONCLUSION

This paper has presented an interpretable machine learning framework for predicting the remaining useful life of rolling bearings using a single accelerometer channel and a compact set of statistical features. The approach combines:

- a Random Forest classifier for robust detection of faulty operating states;
- XGBoost and LSTM regressors for RUL estimation;
- SHAP based analysis for model interpretability.

On the NASA Bearing dataset, the fault classifier achieves 98.6% accuracy and the hybrid RUL estimator improves MAE/RMSE by approximately 15–20% compared to the individual regressors. The SHAP analysis highlights RMS vibration, crest factor and dominant spectral components as key predictors of degradation, in agreement with engineering intuition.

Because the proposed pipeline relies on a single vibration channel, simple features and relatively lightweight models, it is well suited for deployment in resource constrained industrial monitoring systems.

Reproducibility. To facilitate independent verification, we report the full feature list, preprocessing settings, and model configurations. The code for data preparation and model training, together with the train/test split specification and random seeds, is available from the corresponding author upon reasonable request.

Future work will focus on:

- extending the approach to multi sensor configurations and variable operating conditions;
- integrating physics informed constraints and uncertainty quantification into the RUL models;
- validating the framework on real industrial datasets from different sectors (e.g., pumps, compressors, wind turbines), and exploring online learning to adapt to evolving conditions.

CONFLICT OF INTEREST: The authors declare no conflict of interest.

FUNDING: The article was prepared within the framework of the project IRN BR24992785 ‘Organization and conduct of comprehensive research to ensure a sustainable agro-industrial complex of the Kostanay region by creating a research and technological center’.

INSTITUTIONAL REVIEW BOARD STATEMENT: This study complied with applicable institutional and national ethical standards. As the research was purely technical and did not involve human participants or animals, the requirement for Institutional Review Board approval is Not applicable.

INFORMED CONSENT STATEMENT: Not applicable.

DATA AVAILABILITY STATEMENT: The data supporting the findings of this technical study are available from the corresponding author upon reasonable request.

ACKNOWLEDGEMENTS: The authors express their gratitude to the anonymous reviewers for their valuable comments, which helped improve the quality of the article.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES: During the preparation of this work, the authors did not use any artificial intelligence (AI) or AI-assisted technologies in the writing process.

REFERENCES

Shen, S., et al. (2021). A physics-informed deep learning approach for bearing fault detection. Engineering Applications of Artificial Intelligence, 103, 104295. https://doi.org/10.1016/j.engappai.2021.104295

Sanakkayala, D. C., et al. (2022). Explainable AI for bearing fault prognosis using deep learning techniques. Micromachines, 13(9), 1471. https://doi.org/10.3390/mi13091471

Li, J., Liu, J., & Ren, B. (2024). The impact of random parameter distribution on RVFL model performance in bearing fault diagnosis. International Journal of Machine Learning and Cybernetics, 16(12), 9773–9785. https://doi.org/10.1007/s13042-024-02319-9

Neto, D., Petrella, L., et al. (2023). A hybrid deep learning-based approach for rolling bearing fault prognostics. IFAC-PapersOnLine, 56(2), 6588–6593. https://doi.org/10.1016/j.ifacol.2023.10.311

Nalbantov, G., et al. (2021). Predictive maintenance tool for non-intrusive inspection systems. arXiv preprint. https://doi.org/10.48550/arXiv.2103.01044

Sun, B., et al. (2025). Remaining useful life prediction of rolling bearings based on CBAM-CNN-LSTM. Sensors, 25(2), 554. https://doi.org/10.3390/s25020554

Lee, J., Qiu, H., Yu, G., Lin, J., & Rexnord Technical Services. (2007). Bearing data set [Data set]. NASA Ames Prognostics Data Repository. https://data.nasa.gov/dataset/ims-bearings

Койбагаров, М. К., Исабеков, Ж. Н., Курмангалиева, Л. А., Байтурганова, В. К., & Рахметова, П. М. (2025). Разработка системы предиктивного технического обслуживания на основе машинного обучения. Вестник Университета Шакарима. Серия технические науки, 2(18), 121–128. https://doi.org/10.53360/2788-7995-2025-2(18)-14 // Koibagarov, M. K., Isabekov, Zh. N., Kurmangalieva, L. A., Baiturganova, V. K., & Rakhmetova, P. M. (2025). Razrabotka sistemy prediktivnogo tekhnicheskogo obsluzhivaniya na osnove mashinnogo obucheniya [Development of a predictive maintenance system based on machine learning]. Vestnik Universiteta Shakarima. Seriya tekhnicheskie nauki, 2(18), 121–128. https://doi.org/10.53360/2788-7995-2025-2(18)-14 (In Russ.)

Байгожанова, Д. С., Нартай, К. К., Ахитова, Р. С., Кемель, А. М., & Ибраева, Н. Н. (2025). Проектирование и реализация интеллектуальной системы управления на основе машинного обучения. Вестник КазАТК, 3(138), 414–426. https://doi.org/10.52167/1609-1817-2025-138-3-414-426 // Baigozhanova, D. S., Nartai, K. K., Akhitova, R. S., Kemel, A. M., & Ibraeva, N. N. (2025). Proektirovanie i realizatsiya intellektualnoi sistemy upravleniya na osnove mashinnogo obucheniya [Design and implementation of an intelligent control system based on machine learning]. Vestnik KazATK, 3(138), 414–426. https://doi.org/10.52167/1609-1817-2025-138-3-414-426 (In Russ.)

Сагатова, Л. Б., Бортебаев, С. А., Колга, А. Д., Елемесов, К. К., & Басканбаева, Д. Д. (2025). Оценка технического состояния центробежных насосных агрегатов по вибрационным параметрам. Наука и техника Казахстана, 3, 135–153. https://doi.org/10.48081/QGAN2347 // Sagatova, L. B., Bortebaev, S. A., Kolga, A. D., Elemesov, K. K., & Baskanbaeva, D. D. (2025). Otsenka tekhnicheskogo sostoyaniya tsentrobezhnykh nasosnykh agregatov po vibratsionnym parametram [Assessment of the technical condition of centrifugal pumping units by vibration parameters]. Nauka i tekhnika Kazakhstana, 3, 135–153. https://doi.org/10.48081/QGAN2347 (In Russ.)

Hu, J., & Szymczak, S. (2023). A review on longitudinal data analysis with random forest. Briefings in Bioinformatics, 24(2), bbad002. https://doi.org/10.1093/bib/bbad002

Wiens, M., Verone-Boyle, A., Henscheid, N., Podichetty, J. T., & Burton, J. (2025). A tutorial and use case example of the eXtreme Gradient Boosting (XGBoost) artificial intelligence algorithm for drug development applications. Clinical and Translational Science, 18(3), e70172. https://doi.org/10.1111/cts.70172

Krichen, M., & Mihoub, A. (2025). Long short-term memory networks: A comprehensive survey. AI, 6(9), 215. https://doi.org/10.3390/ai6090215

Ponce-Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. Clinical and Translational Science, 17(11), e70056. https://doi.org/10.1111/cts.70056

Мадин, В. А., Комаров, Д. Н., Салыкова, О. С., Салыков, Б. Р., Зарубин, М. Ю., & Кобжасаров, Т. Ж. (2025). Предиктивное обслуживание шпиндельных подшипников на основе вибродиагностики на примере токарного станка СКЕ 6136 6150. Вестник КазУТБ, 3(28), 63–79. https://vestnik.kaztbu.edu.kz/index.php/kazutb/article/view/873/544 // Madin, V. A., Komarov, D. N., Salykova, O. S., Salykov, B. R., Zarubin, M. Yu., & Kobzhasarov, T. Zh. (2025). Prediktivnoe obsluzhivanie shpindelnykh podshipnikov na osnove vibrodiagnostiki na primere tokarnogo stanka CKE 6136 6150 [Predictive maintenance of spindle bearings based on vibration diagnostics: The case of the CKE 6136 6150 lathe]. Vestnik KazUTB, 3(28), 63–79. https://vestnik.kaztbu.edu.kz/index.php/kazutb/article/view/873/544 (In Russ.)

Авторлар туралы мәліметтер
Информация об авторах
Information about authors

Жарлыкасова Анара Нурлановна – докторант, Ахмет Байтұрсынұлы атындағы Қостанай өңірлік университеті, Қостанай қ., Қазақстан
Жарлыкасова Анара Нурлановна – докторант, Костанайский Региональный университет имени Ахмет Байтұрсынұлы, г. Костанай, Казахстан
Zharlykassova Anara Nurlanovna – doctoral student, NPJSC Akhmet Baitursynuly Kostanay Regional University, Kostanay, Kazakhstan,
e-mail: anara2719@mail.ru,
ORCID: https://orcid.org/0009-0002-7862-4182,

Салыкова Ольга Сергеевна – техника ғылымдарының кандидаты, Ахмет Байтұрсынұлы атындағы Қостанай өңірлік университеті, Қостанай қ., Қазақстан
Салыкова Ольга Сергеевна – кандидат технических наук, Костанайский Региональный университет имени Ахмет Байтұрсынұлы, г. Костанай, Казахстан
Salykova Olga Sergeevna – Candidate of Technical Sciences, NPJSC Akhmet Baitursynuly Kostanay Regional University, Kostanay, Kazakhstan,
e-mail: solga0603@mail.ru,
ORCID: https://orcid.org/0000-0002-8681-4552

Маусымбаева Самал Батырбековна – жаратылыстану ғылымдарының магистрі, Ахмет Байтұрсынұлы атындағы Қостанай өңірлік университеті, Қостанай қ., Қазақстан
Маусымбаева Самал Батырбековна – магистр естественных наук, Костанайский Региональный университет имени Ахмет Байтұрсынұлы, г. Костанай, Казахстан
Maussymbaeva Samal Batyrbekovna – Master of Technical Sciences, NPJSC Akhmet Baitursynuly Kostanay Regional University, Kostanay, Kazakhstan,
e-mail: msamal_93@mail.ru,
ORCID: https://orcid.org/0009-0006-3407-7702