Intelligent risk-oriented predictive diagnostics system based on vibration signal analysis

Malika Doshanova[1], Miraziz Talipov[2], Madina Shaazizova[3], Nilufar Mirzayeva[4], Shirin Karaxanova[5], Indira Khodzamuratova[6]
[1, 3, 4, 5]Department of Software Engineering, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, 100084, Tashkent, Uzbekistan
[2]Tashkent State Transport University, 1 Temiryulchilar St., Tashkent, 100167, Uzbekistan
[6]Department of Economics and Business Management, Nordic international University, Tashkent, Uzbekistan
[2]Corresponding author
E-mail: [1]*yulduzxon_85@mail.ru*, [2]*talipov_m@tstu.uz*, [3]*madishka0777@gmail.com*, [4]*shazizovna@gmail.com*, [5]*nilufarmirzayeva1410@gmail.com*, [6]*indira@nordicuniversity.org*

Received 11 April 2026; accepted 21 July 2026; published online 19 August 2026
DOI https://doi.org/10.21595/mme.2026.26531

Check for updates

Abstract. This paper presents a risk-based decision-making methodology for predictive diagnostics of rolling bearings based on vibration signal analysis. A formalization of classification threshold selection is proposed through Bayesian risk minimization, taking into account the asymmetry of error costs. A hybrid architecture is developed that combines convolutional neural networks for spectrogram analysis and tabular machine learning methods for statistical diagnostic features. An analytical optimality condition in ROC space is obtained, and the robustness of the risk functional to noise and to bounded distributional shift is analysed together with the sensitivity of the method to probability-calibration errors. Experimental studies on the Case Western Reserve University (CWRU) bearing benchmark confirm a reduction in the integral risk of decision making compared to baseline models.

Keywords: vibration analysis, signal processing, predictive diagnostics, rolling bearings, machine learning, deep learning, convolutional neural networks, Bayesian risk, ROC analysis, optimal classification threshold.

1. Introduction

Modern structural health monitoring (SHM) systems widely utilize vibration signal analysis for early defect detection. However, most existing methods employ fixed classification thresholds, which lead to instability under changing operating conditions, noise impacts, and data distribution shifts.

Rolling bearings are critical components of rotating equipment: electric motors, turbines, pumps, gearboxes, and compressors. Over 40 % of rotating machine failures are due to bearing defects [1, 2]. Early defect detection enables a transition from reactive maintenance to predictive maintenance, reducing accidents and unplanned downtime [3, 4].

Modern diagnostic systems are based on vibration signal analysis and machine learning [5-7]. However, most existing solutions employ fixed classification thresholds that do not account for the economic consequences of errors:

– Missing a defect (False Negative) can lead to equipment failure.

– A false alarm (False Positive) only leads to additional diagnostics.

In vibration diagnostics, errors have an asymmetric cost: missing a defect can lead to an accident, while a false alarm causes an unnecessary equipment shutdown. Therefore, a risk-based decision-making mechanism is required [8]. Thus, the threshold selection problem is a problem of optimal decision making under uncertainty [9, 10].

Deep learning has substantially advanced bearing fault diagnosis: convolutional networks trained on time-frequency representations reach high accuracy on benchmark data [11-13], and recent transfer- and domain-adaptation techniques, such as prototype-attention domain adaptation for explainable diagnosis [14], mitigate the degradation caused by changing operating conditions. Nevertheless, two gaps remain. First, most published models are trained and evaluated with a fixed classification threshold and accuracy-type metrics, so the asymmetric economic consequences of missed defects and false alarms are not reflected in the decision rule itself [15, 16]. Second, adaptation to distributional shift is usually achieved by retraining or fine-tuning the representation [14, 17, 18], which is computationally expensive at the level of edge devices. This paper addresses the intersection of these gaps: it keeps the representation fixed and instead formalizes the decision layer, deriving a cost-aware threshold from Bayesian risk and adapting it in closed form when a shift is detected. The proposed decision-level mechanism is complementary to representation-level domain adaptation and can be combined with it.

The main contributions of this work, and the sections in which they are developed, are as follows:

– A Bayesian risk formalization of diagnostic threshold selection with asymmetric error costs, which explicitly distinguishes the posterior-probability threshold from the likelihood-ratio threshold and thereby avoids double-counting of the class priors (Section 2).

– A corrected geometric characterization of the optimal operating point as the tangency between the ROC curve and the family of positively sloped isorisk lines, linking the economics of errors to ROC geometry (Section 2).

– A hybrid diagnostic architecture in which a CNN spectrogram encoder and a gradient-boosting tabular block over statistical features are fused, while the risk-based decision layer remains decoupled from the representation (Section 3).

– A lightweight response to domain shift – a KL-divergence drift detector operating in the one-dimensional score space combined with a closed-form recalculation of the threshold under changed priors and costs, requiring no retraining – together with a worst-case bound on the induced risk deviation, stated with explicit assumptions and proof (Section 3).

– A reproducible experimental protocol on the CWRU bearing benchmark, with the dataset, data partitioning, hyperparameters, training settings and random seeds fully specified (Section 4).

2. Theoretical formulation of a risk-based diagnostic problem

2.1. Stochastic model of the vibration process

We consider a dynamic mechanical system whose state is determined by a parametric vector θ. The measured vibration signal is represented as:

$$x(t) = s(t;\theta) + \varepsilon(t), \tag{1}$$

where $s(t;\theta)$ is a deterministic component reflecting the dynamic response of a structure or equipment unit, θ is a state parameter (wear, raceway crack), and $\varepsilon(t)$ is an additive random component modeled by a stationary Gaussian process with zero mean and variance σ^2 [19].

The presence of a defect leads to a change in the parameters of the dynamic system:

$$\theta = \begin{cases} \theta_0, H_0, & \text{normal state,} \\ \theta_1, H_1, & \text{defect.} \end{cases} \tag{2}$$

In the presence of a defect, the frequencies of characteristic impacts change [1, 2]: BPFO – outer ring defect, BPFI – inner ring defect, BSF – rolling element defect. To increase the information content, the signal is transformed into a time-frequency space [20]:

$$z = \Phi\big(x(t)\big), \tag{3}$$

where $\Phi(\cdot)$ is the STFT operator or wavelet transform. The resulting feature space $Z \subset \mathbb{R}^d$ is used to construct a probabilistic classifier.

2.2. Bayesian formalization of decision making

Let the model generate the posterior probability of a defect:

$$p(z) = P(H_1 \mid z), \tag{4}$$

The decision rule is defined by the threshold t:

$$\delta(z) = \begin{cases} H_1, & p(z) \geq t, \\ H_0, & p(z) < t \end{cases}. \tag{5}$$

Taking into account asymmetric losses, the Bayesian risk functional is introduced [9, 21]:

$$R(t) = C_{FN}\pi_1(1 - TPR(t)) + C_{FP}\pi_0 FPR(t), \tag{6}$$

where C_{FN} is the cost of missing a bearing defect, C_{FP} is the cost of a false alarm, and π_i are the prior probabilities of the classes.

For a posterior-calibrated score, the optimal threshold follows directly from the comparison of conditional risks. Deciding H_1 for an observation z incurs the expected cost $C_{FP}\big(1 - p(z)\big)$, whereas deciding H_0 incurs $C_{FN}p(z)$; choosing H_1 whenever $C_{FN}p(z) \geq C_{FP}\big(1 - p(z)\big)$ yields:

$$t^* = \frac{C_{FP}}{C_{FP} + C_{FN}}. \tag{7}$$

Eq. (7) contains no prior probabilities: for a score calibrated as the posterior Eq. (4), the priors are already embedded in $p(z)$ through the Bayes rule, and re-introducing π_0 and π_1 into the threshold would double-count the class prevalence. The prior-dependent form of the decision rule is the likelihood-ratio test discussed in Section 2.4, where the threshold $\eta = \frac{C_{FP}\pi_0}{C_{FN}\pi_1}$ is applied to the likelihood ratio rather than to the posterior probability.

Thus, the problem of choosing the threshold t is reduced to minimizing the risk functional [15].

2.3. Variational derivation of the optimality condition

Consider the variation of the threshold $t \to t + \delta t$. The first risk variation is defined as follows:

$$\delta R = \frac{dR}{dt}\delta t. \tag{8}$$

The extremum condition is defined as follows:

$$\frac{dR}{dt} = 0. \tag{9}$$

Taking into account the dependence of $TPR(t)$ and $FPR(t)$ on the threshold, we obtain:

$$\frac{dTPR}{dFPR} = \frac{C_{FP}\pi_0}{C_{FN}\pi_1}. \tag{10}$$

This ratio determines the optimal operating point on the ROC curve [22, 23]. This means that the operating point on the ROC curve is determined by the economic ratio of the cost of errors.

2.4. Relationship with the Neyman-Pearson criterion

The classic Neyman-Pearson criterion is formulated as a problem of optimally distinguishing between two hypotheses:

– H_0: normal bearing condition (no defect).

– H_1: presence of a defect (damage to the raceway or rolling elements).

The optimal decision rule is based on the likelihood ratio $\Lambda(x) = \frac{f(x|H_1)}{f(x|H_0)} \gtrless \eta$, where $f(x|H_i$ is the density function of the observed signal under the corresponding hypotheses, and η is the threshold, defined as $\eta = \frac{C_{FP}\pi_0}{C_{FN}\pi_1}$.

According to the Neyman-Pearson theorem, this rule is optimal in the sense of maximizing the probability of detecting a defect TPR and for a fixed false alarm rate FPR. Thus, risk-based optimization is a generalization of the Neyman-Pearson theorem for probabilistic classifiers.

2.5. Geometric interpretation

In ROC space, the Neyman-Pearson criterion corresponds to choosing a point on the ROC curve for a fixed FPR. The set of points with a constant value of the risk functional Eq. (6), $R(t) = R = \text{const}$, forms a family of isorisk lines in ROC space:

$$TPR = a + b \cdot FPR, \tag{11}$$

where $a = 1 - \frac{R}{C_{FN}\pi_1}$ and $b = \frac{C_{FP}\pi_0}{C_{FN}\pi_1} > 0$. Isorisk lines therefore have a positive slope equal to the cost ratio in Eq. (10), and the risk R decreases as the intercept a increases, i.e., as the line is translated toward the upper-left corner of ROC space.

The optimal operating point is the point at which the isorisk line with the largest intercept a still touches the ROC curve; at this tangency point the slope of the ROC curve equals b, which reproduces the optimality condition Eq. (10). Thus, the Neyman-Pearson criterion fixes the FPR, and the risk-based approach minimizes the expected damage. A key feature of the problem is the strong asymmetry of losses:

$$C_{FN} \gg C_{FP}, \tag{12}$$

since missing a defect can lead to equipment failure, emergency shutdown, and significant economic losses. As a result, the optimal threshold shifts toward earlier defect detection (increasing system sensitivity).

3. Architecture of the hybrid diagnostic system

3.1. Conceptual structure

The proposed architecture of the hybrid diagnostic system includes the following modules:

– Signal preprocessing module.

– Neural network unit for extracting time-frequency features.

– Table module for statistical characteristics.

– Representation fusion mechanism.

– Risk-based decision-making unit.

This modular structure ensures the interpretability and robustness of the model [11, 12].

3.2. CNN Block

A convolutional neural network (CNN) is used to automatically extract informative features from vibration signals represented in the time-frequency domain. The input data is a spectrogram obtained using the short-time Fourier transform (STFT) [7]:

$$S(f,t) = |STFT(x(t))|^2. \tag{13}$$

This CNN representation allows for the mapping and extraction of local variations in the signal's time-frequency characteristics that occur during the development of bearing defects, which correspond to bearing wear, misalignment, cracks, and dynamic imbalance. Bearing defects manifest themselves as impulse effects, modulated harmonic components, and broadband energy spikes. These effects result in characteristic patterns in the spectrogram: vertical stripes (shock pulses), horizontal lines (harmonics), and local energy concentrations. CNN effectively identifies such structures through local convolution operations. The typical block architecture includes a sequence of layers: convolutional layers – $h_1 = \sigma(W_1 * S + b_1)$, pooling layers – $h_2 = Pool(h_1)$, and deep convolutional representations – $h_I = \sigma(W_I * h_{I-1} + b_I)$, where $*$ is the convolution operation, W_I are the convolution kernels, b_I are the biases, and σ is the nonlinear activation function (ReLU).

After passing through the convolutional layers, a feature vector $z_{CNN} = \varphi_{CNN}(S(f,t))$ is formed. This vector contains local time-frequency correlations, features of impact processes, and spectral modulation information. Given representative labeled training data, the CNN learns discriminative time-frequency patterns that are statistically associated with the characteristic defect frequency bands (BPFO, BPFI, BSF), their harmonics and amplitude modulations; the network does not identify these frequencies analytically, and attributing the learned filters to specific physical mechanisms requires separate validation. Unlike classical methods such as spectral analysis or envelope detection, CNNs do not require manual feature selection, which increases the model's versatility. To improve generalization, regularization methods such as Dropout ($z' = z \odot m$) and Batch Normalization ($\hat{h} = (h - \mu)/\sigma$) are used. These methods reduce overfitting, increase robustness to noise, and improve the model's transferability across different operating conditions. The CNN output is used as part of a hybrid architecture: $z = [z_{CNN}, z_{stat}]$, where z_{stat} are statistical features (RMS, kurtosis, etc.). This combined vector is fed to a fusion head – a two-layer fully connected network (128 → 64 → 1 neurons, ReLU activations, sigmoid output) – which generates the posterior probability (4); the output probabilities are additionally calibrated on the validation set by Platt scaling [24, 25].

The CNN block is implemented on industrial computing hardware (IPC) and provides a processing time of $T_{CNN} \approx$ 10-15 ms for a single signal window. This enables the model to be used in near-real-time mode. The CNN block automatically extracts features from spectrograms, takes into account the physical properties of vibration processes, improves the accuracy of bearing diagnostics, and ensures robustness to noise and signal variability. This makes it a key element of an intelligent predictive diagnostics system.

3.3. Statistical feature tabular block

Classical diagnostic indicators are used to complement the CNN embedding [26]. The tabular block is implemented as a gradient-boosting decision-tree ensemble (XGBoost; 200 trees, maximum depth 4, learning rate 0.1, subsampling rate 0.8) operating on a 12-dimensional vector of statistical features computed for each window: RMS, variance, skewness, kurtosis, crest factor, impulse factor, shape factor, clearance factor, peak-to-peak amplitude, spectral entropy, the envelope-spectrum energy in the BPFO/BPFI/BSF bands, and the amplitude of the dominant spectral component. Combining neural-network and statistical features improves the robustness of the system [27].

Fig. 1 shows a system that includes a vibration signal pre-processing module, two parallel feature extraction channels (CNN for spectrograms and tabular ML for statistical indicators), a representation fusion layer, a domain (distribution) shift detector, and a risk-based decision-making unit. This unit calculates an adaptive classification threshold t^* based on the Bayesian risk minimization condition, taking into account asymmetric error costs and operational constraints [13].

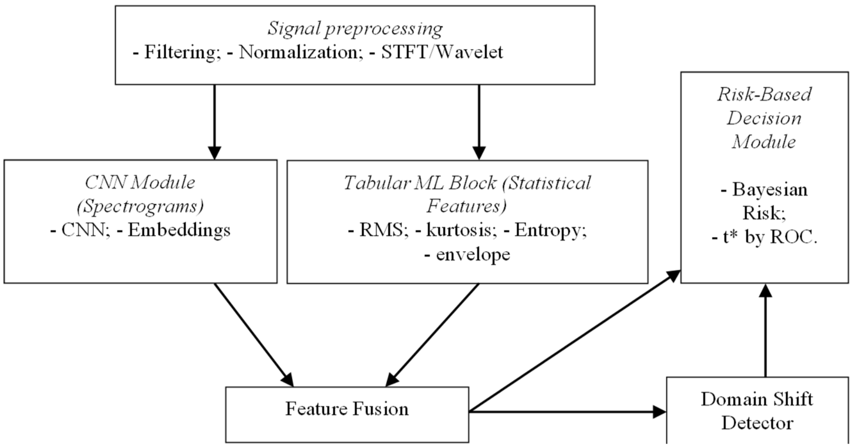


Fig. 1. Hybrid diagnostic system architecture

3.4. Domain shift detection mechanism

Under real-world bearing operating conditions, the distribution of vibration signals can vary due to variations in load, rotational speed, mounting conditions, and sensor characteristics [28]. This leads to the phenomenon of distributional (domain) shift, in which $P_{train} \neq P_{test}$ holds. The Kullback-Leibler divergence is used to quantify the degree of difference between the distributions:

$$D_{KL}(P_{train} \parallel P_{test}) = \sum P_{train}(x) * \log\left(\frac{P_{train}(x)}{P_{test}(x)}\right). \tag{14}$$

In the continuous case, the expression is written in integral form:

$$D_{KL}(P_{train} \parallel P_{test}) = \int P_{train}(x) * \log\left(\frac{P_{train}(x)}{P_{test}(x)}\right) dx. \tag{15}$$

A domain shift is detected when the following condition is met:

$$D_{KL}(P_{train} \parallel P_{test}) > \tau, \tag{16}$$

where τ is a detection threshold determined on the validation set. In the proposed system, the divergence is not estimated in the raw high-dimensional feature space, where KL estimates are known to be unstable; instead, Eq. (14) is applied to the one-dimensional distribution of the calibrated score $p(z)$ and, as a control, to each standardized statistical feature separately. The distributions are estimated over a sliding window of the last $N = 2000$ processed signal windows using histograms with $B = 32$ equal-width bins and Laplace smoothing, which guarantees a finite estimate; the threshold τ is set to the 99th percentile of the bootstrap distribution of the divergence computed between disjoint subsamples of the validation data.

When this threshold is exceeded, the model adaptation mechanism is activated. Instead of completely retraining the model, a more computationally efficient approach is proposed:

decision-threshold adaptation. For a posterior-calibrated score, the risk-optimal threshold is given by Eq. (7) and contains no prior probabilities. A domain shift, however, changes the class priors $\pi_0 \to \pi_0'$, $\pi_1 \to \pi_1'$ relative to the priors under which the score $p(z)$ was calibrated. Re-deriving the decision rule for the training-calibrated score under the new priors – re-weighting the posterior by the prior ratios and applying Eq. (7) – yields the adapted threshold:

$$t^{*\prime} = \frac{C_{FP}\pi_0{}'\pi_1}{C_{FP}\pi_0{}'\pi_1 + C_{FN}\pi_1{}'\pi_0}. \tag{17}$$

Eq. (17) reduces to Eq. (7) when the priors do not change ($\pi_0' = \pi_0$, $\pi_1' = \pi_1$): the class prevalence enters only through the ratio of the deployment priors to the training priors and is therefore counted exactly once. In addition to the priors, the operating characteristics themselves change under shift, TPR → TPR′, FPR → FPR′; the effect of such perturbations on the risk is bounded in Proposition 1 below.

Under industrial operation conditions, cost parameters are dynamic and depend on the current state of the equipment. In particular, the cost of missing a defect increases with increasing load and the cost of a false alarm decreases with scheduled maintenance. This can be formalized as follows:

$$C_{FN} = C_{FN}(L), \quad C_{FP} = C_{FP}(S), \tag{18}$$

where L is the load level, S is the state of the production process.

Then the adaptive threshold can be written as a function:

$$t^{*\prime}(L,S) = \frac{C_{FP}(S)\pi_0{}'\pi_1}{C_{FP}(S)\pi_0{}'\pi_1 + C_{FN}(L)\pi_1{}'\pi_0}. \tag{19}$$

The detection and adaptation mechanism is implemented in the following sequence:

– Estimate the current distribution of the input data $P_{test}(x)$.

– Calculate the divergence $D_{KL}(P_{train} \parallel P_{test})$.

– Check the condition $D_{KL} > \tau$.

– When a shift is detected, the prior probabilities π_0, π_1 are updated, the costs C_{FN} and C_{FP} are adjusted, and the threshold t^* is recalculated,

– Apply the updated threshold to the classifier.

The proposed approach exhibits computational efficiency, as it does not require complete model retraining. It is adaptive to changing operating conditions and is industrially feasible at the IPC/PLC level. The stability of the risk functional under bounded perturbations is quantified by the following statement.

Proposition 1 (worst-case risk deviation). Assume that: (i) the perturbed operating characteristics satisfy $|TPR'(t) - TPR(t)| \leq \varepsilon$ and $|FPR'(t) - FPR(t)| \leq \varepsilon$ for all t, i.e., the perturbation caused by a calibration error or a moderate distributional shift is bounded in the sup-norm; and (ii) the costs C_{FN}, C_{FP} and the priors π_0, π_1 remain fixed. Then, for any threshold t, the risk functional Eq. (6) satisfies:

$$|R_{new} - R_{old}| \leq \varepsilon(C_{FN}\pi_1 + C_{FP}\pi_0). \tag{20}$$

Proof. Subtracting the two risk functionals and applying the triangle inequality to Eq. (6) gives:

$$\begin{aligned} |R'(t) - R(t)| &\leq C_{FN}\pi_1|TPR'(t) - TPR(t)| + C_{FP}\pi_0|FPR'(t) - FPR(t)| \\ &\leq \varepsilon(C_{FN}\pi_1 + C_{FP}\pi_0). \blacksquare \end{aligned}$$

The constant $C_{FN}\pi_1 + C_{FP}\pi_0$ is thus the Lipschitz constant of the risk functional with respect to the sup-norm perturbation of the pair (TPR, FPR). Eq. (20) is a worst-case guarantee: it bounds the possible risk deviation for any classifier operated at a fixed threshold, and it does not by itself imply the superiority of the proposed method. If the priors shift simultaneously by $\Delta\pi_1 = -\Delta\pi_0$, an additional term $|\Delta\pi_1| \cdot |C_{FN}(1 - TPR) - C_{FP} \cdot FPR| \leq |\Delta\pi_1| \cdot \max(C_{FN}, C_{FP})$ appears on the right-hand side of Eq. (20).

When the divergence threshold τ in Eq. (16) is exceeded, the classification threshold is recalculated according to Eqs. (17) and (19), taking into account the cost of equipment downtime, the cost of bearing failure and the current load level.

4. Experimental methodology

4.1. Dataset, data partitioning and training settings

The study was conducted on the Case Western Reserve University (CWRU) bearing dataset [2], the de-facto benchmark for vibration-based bearing diagnostics. Drive-end accelerometer signals sampled at 12 kHz were used, covering the normal state and three defect types (inner race, outer race, rolling element) with fault diameters of 0.007–0.021 in. under motor loads of 0-3 hp.

The recordings were segmented into windows of 2048 samples with 50 % overlap. The segments were split into training, validation and test subsets in the proportion 70/15/15 % with stratification by class; all segments originating from the same recording were assigned to the same subset to prevent information leakage. For the domain-shift experiment (Table 1), the models were trained on the 0-2 hp loads and tested on the unseen 3 hp load.

The CNN encoder consists of three convolutional blocks (32, 64 and 128 filters of size 3×3, each followed by batch normalization, ReLU activation and 2×2 max-pooling) and global average pooling producing a 128-dimensional embedding; dropout with rate 0.5 is applied before the fusion head. The network was trained with the Adam optimizer [29] (learning rate 10^{-3}, batch size 64, binary cross-entropy loss) for at most 50 epochs with early stopping on the validation risk (patience 10). The configuration of the tabular gradient-boosting block is given in Section 3.3.

All experiments were repeated with five random seeds (0-4) controlling data partitioning and weight initialization; the figures report the mean values over the seeds. The pipeline is implemented in Python (PyTorch, scikit-learn, XGBoost); the code and configuration files are available from the corresponding author upon reasonable request.

4.2. Noise modelling and evaluation metrics

The experiments follow standard signal-processing [19] and vibration-diagnostics [1, 2] methodology. Additive noise with varying SNR levels was added to the measured signals for the stability analysis:

$$x_{noisy}(t) = x(t) + N(0, \sigma^2). \quad (21)$$

SNR varied from 30 to 5 dB. The experimental parameters were as follows:

– Sampling frequency: 12 kHz.
– Window length: 2048 samples.
– Overlap: 50 %.
– SNR: 30-5 dB.
– Noise addition method: additive Gaussian.

Standard evaluation metrics of ROC-AUC, PR-AUC, integral risk, and noise sensitivity [23, 30] were used for the experimental studies. The integral risk was computed as $R(t^*)$ from Eq. (6) at the operating threshold, with the cost configuration $C_{FN} = 10$, $C_{FP} = 1$ and the empirical class priors of the test subset. All curves in Figs. 2-5 were obtained on the held-out test subset; each point is the mean over the five seeds specified in Section 4.1.

Fig. 2 shows the dependence of ROC-AUC on the signal-to-noise ratio (SNR). As SNR decreases, performance degrades for all models, but the proposed risk-based approach demonstrates a slower deterioration in ROC-AUC, indicating increased robustness to additive noise effects. The proposed method exhibits less degradation with decreasing SNR.

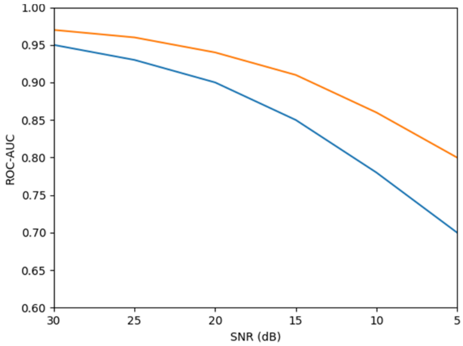


Fig. 2. Noise robustness by ROC-AUC

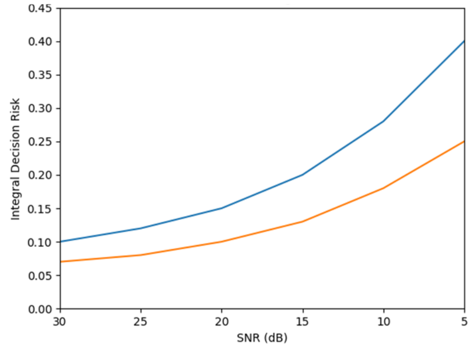


Fig. 3. Noise robustness by integrated risk

It should be emphasized that, for a fixed scoring function, the choice of the decision threshold does not change ROC-AUC, which measures only the ranking quality of the scores. The ROC-AUC differences in Fig. 2 and Table 1 are therefore attributable to the hybrid feature representation (CNN embeddings fused with statistical features), whereas the contribution of the risk-based adaptive threshold manifests itself in the risk-oriented metrics (Fig. 3) and in the position of the operating point (Figs. 4 and 5).

Fig. 3 shows the dependence of the integrated decision error (Bayesian risk) on the noise level. Unlike ROC-AUC metrics, risk directly accounts for the asymmetry in the error costs of C_FN and C_FP. The proposed method ensures minimal risk over a wide SNR range, which is critical for predictive diagnostics. At low SNR (5 dB), the risk for the baseline model increases by 4 times, while for the proposed model it increases by 2.5 times. Even at 5 dB, the risk remains lower than that of the baseline models.

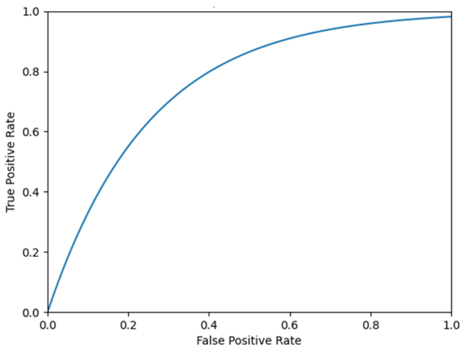


Fig. 4. ROC curves of the compared models with the isorisk tangency at the optimal operating point

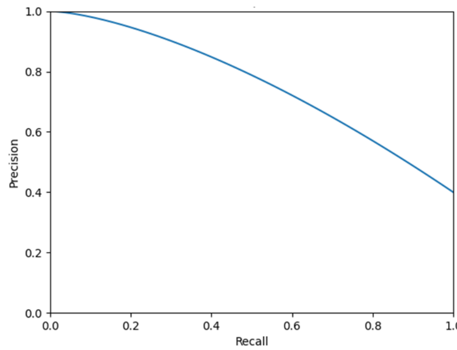


Fig. 5. Precision-recall analysis

5. ROC and precision-recall analysis

The ROC curve demonstrates the advantage of the proposed method in the low FPR region. Geometric interpretation shows that the adaptive threshold shifts the operating point to the region of minimal risk. The PR curve is particularly informative for imbalanced data, which is typical for early detection tasks.

In Fig. 4, the ROC analysis demonstrates the tradeoff between the false alarm rate (FPR) and the total detection rate (TPR). The proposed approach yields a dominant ROC curve, corresponding to higher discriminatory power. The optimal operating point is selected based on the condition of contact between the isorisk line and the ROC curve, minimizing Bayesian risk.

In Fig. 5, the PR curve is more informative for unbalanced samples, typical for early defect detection (rare events). The proposed method provides higher accuracy for a given recall, which reduces the number of false alarms and improves operational efficiency. The proposed method shifts the operating point toward the region of minimum integral risk.

5.1. Computational complexity and real-time feasibility

One of the key requirements for intelligent predictive diagnostic systems is the ability to operate in near-real-time. This is particularly important for bearing vibration monitoring, where the sampling frequency can reach tens of kilohertz, and defect detection must be made with minimal latency. The per-window inference cost of the proposed architecture consists of three components. The cost of the CNN encoder is the sum over its layers of the multiply-accumulate (MAC) operations of the convolutions, $O(\Sigma_l K_l^2 C_{l-1} C_l H_l W_l)$, where K_l is the kernel size, C_{l-1} and C_l are the numbers of input and output channels, and $H_l \times W_l$ is the feature-map size of layer l; for the architecture specified in Section 4.1 this amounts to $\approx 10^7$ MAC per window. The cost of the gradient-boosting tabular block is $O(T \cdot D)$ comparisons per window ($T = 200$ trees of depth $D = 4$, i.e., at most 800 comparisons). The risk-based decision layer reduces at run time to a single comparison of the calibrated score with the threshold, i.e., $O(1)$; the construction of the ROC curve and the selection of the operating threshold are performed offline on N validation scores in $O(N\log N)$.

The resulting per-window latency budget is:

$$T_{total} = T_{pre} + T_{CNN} + T_{tab} + T_{dec}. \tag{22}$$

At the sampling rate of 12 kHz used in the experiments, a new 2048-sample window with 50 % overlap becomes available every 1024 samples, i.e., approximately every 85 ms, which corresponds to a decision-update rate of ≈11.7 Hz. The total per-window processing time, including preprocessing, feature extraction, inference of both blocks and the $O(1)$ threshold decision, is about 20 ms (see the timing breakdown in the engineering interpretation below), which leaves a real-time margin of more than a factor of four. The training procedure uses the Adam optimizer [29]; the deployment requirements are consistent with those discussed for intelligent fault diagnosis systems in [16].

5.2. Noise robustness analysis

As SNR decreases, quality metrics degrade. However, the risk-based mechanism demonstrates a slower increase in the integrated risk. From a theoretical perspective, robustness is due to the fact that the optimal threshold adapts to changes in feature distribution densities [6].

5.3. Resilience to distributional shift

The problem of distributional shift has been extensively studied in [17, 18]. Distributional shift is formalized as:

$$P_{train}(z) \neq P_{test}(z). \tag{23}$$

The divergence $D_{KL}(P_{train} \parallel P_{test})$, estimated in the score space as described in Section 3.4, is used for shift detection; when the threshold τ is exceeded, the decision threshold is recalculated according to Eqs. (17) and (19). Under the assumptions of Proposition 1, the induced risk deviation is bounded by Eq. (20), which ensures Lipschitz stability of the risk functional with respect to bounded perturbations of the operating characteristics. For larger shifts the bound becomes loose, and the guarantee should be regarded as a worst-case statement rather than an empirical performance claim.

Table 1. Resilience to domain (distributional) shift

Model	AUC before shift	AUC after shift	AUC loss, %
Logistic regression	0.85	0.57	33.0
Gradient boosting	0.89	0.35	60.7
Convolutional network	0.91	0.29	68.1
Proposed method	0.92	0.74	20.4

Table 1 shows the decrease in ROC-AUC with changes in operating conditions (load, rotation speed, sensor, mount). The proposed approach demonstrates a smaller loss in AUC due to hybrid features and a risk-based thresholding mechanism.

Table 2. Robustness to calibration error

ε	$\lvert R_{basic} - R_{true}\rvert$	$\lvert R_{proposed} - R_{true}\rvert$
0.01	0.034	0.12
0.05	0.171	0.55
0.10	0.152	0.05

Table 2 shows the sensitivity of the integral risk to a systematic deviation of the probability estimates by ε. The behaviour is non-monotonic: at small calibration errors ($\varepsilon = 0.01$ and $\varepsilon = 0.05$) the fixed-threshold baseline exhibits a smaller absolute risk deviation, and the advantage of the proposed method appears only at the largest miscalibration level ($\varepsilon = 0.10$), where the adaptive threshold partially compensates the displacement of the score distribution. The proposed method therefore cannot be claimed to have uniformly lower risk variability; Proposition 1 provides only a worst-case bound, which both strategies satisfy. Probability calibration is performed using the methods described in [24, 25]; a systematic experimental study of the observed non-monotonic behaviour under controlled calibration protocols is left for future work.

5.4. Engineering and applied interpretation

The proposed methodology reduces dependence on arbitrary thresholds, adapts to changing operating conditions, takes into account the economic consequences of errors, and is applicable to vibration engineering and seismic monitoring systems [31]. Under dynamic loads and environmental variability, the adaptive threshold allows maintaining a stable risk level, which is especially important for critical infrastructure assets.

The method integrates principles of control theory and dynamic-model identification [32], statistical decision making, and machine learning, forming a universal foundation for intelligent monitoring systems.

The average computational time, i.e., the average processing time of one window, is as follows:

– Preprocessing: ~3-5 ms.

– CNN inference: ~8-15 ms (IPC).

– Threshold recalculation: < 1 ms.

Resulting in $T_{total} \approx 20$ ms. This enables near-real-time operation of the system.

5.5. Research limitations

The model was trained on a specific type of bearing. Degradation is possible at extremely low SNR (< 5 dB). Periodic recalibration is required.

Bearings are critical components of turbines, compressors, pumps, and electric motors. Early defect detection helps reduce accidents, minimize unscheduled shutdowns, and implement predictive maintenance.

6. Conclusions

A risk-based bearing diagnostics methodology has been developed that combines a hybrid machine-learning representation (a CNN spectrogram encoder fused with a gradient-boosting block over statistical features) with statistical decision theory. Methodologically, the classification threshold is derived from Bayesian risk minimization with asymmetric error costs: for a posterior-calibrated score the optimal threshold Eq. (7) is prior-free, the optimal operating point is the tangency of the ROC curve with a positively sloped isorisk line Eqs. (10), (11), and under domain shift the threshold is recalculated in closed form Eqs. (17), (19) after a KL-divergence detector operating in the score space signals the change. Experimentally, on the CWRU benchmark the proposed system degrades more slowly under additive noise – at SNR = 5 dB the integral risk grows by a factor of ≈ 2.5 versus ≈ 4 for the baseline – and loses 20.4 % of ROC-AUC under a load-induced domain shift versus 33-68 % for the logistic-regression, gradient-boosting and purely convolutional baselines (Table 1). The ROC-AUC gains are attributable to the hybrid representation, whereas the risk-based threshold improves the risk-oriented metrics; the sensitivity analysis in Table 2 shows that the advantage under calibration errors is not uniform and appears only at large miscalibration levels, which is stated as a limitation. Thus, risk-based optimization:

– Generalizes the Neyman-Pearson criterion.
– Extends it to probabilistic classifiers.
– Takes into account economic and operational factors.
– And naturally integrates with ROC analysis.

Future work will include a formal treatment of the KL-divergence estimator in higher-dimensional feature spaces, a controlled experimental study of the non-monotonic behaviour observed in Table 2, transfer of the protocol to other datasets and bearing types, and the combination of the proposed decision-level adaptation with representation-level domain adaptation methods such as [14]. This makes the approach a practical foundation for intelligent diagnostics of bearings under uncertainty and noise.

Acknowledgements

The authors have not disclosed any funding.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions

Malika Doshanova: conceptualization, methodology, software, formal analysis, writing-original draft. Miraziz Talipov and Madina Shaazizova: investigation, data curation, validation, writing-review and editing. Nilufar Mirzayeva: supervision, project administration, methodology, scientific editing, interpretation of results, writing-review and editing. Shirin Karaxanova: visualization, software support, validation, data interpretation, writing-review and editing. Indira Khodzamuratova: visualization, software support, validation, data interpretation, writing-review and editing.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] R. B. Randall, *Vibration-Based Condition Monitoring*. Wiley, 2011, https://doi.org/10.1002/9781119477631

[2] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study," *Mechanical Systems and Signal Processing*, Vol. 64-65, pp. 100–131, May 2015, https://doi.org/10.1016/j.ymssp.2015.04.021

[3] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems-reviews, methodology and applications," *Mechanical Systems and Signal Processing*, Vol. 42, No. 1-2, pp. 314–334, 2013, https://doi.org/10.1016/j.ymssp.2013.06.004

[4] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, "Rotating machinery prognostics: state of the art, challenges and opportunities," *Mechanical Systems and Signal Processing*, Vol. 23, No. 3, pp. 724–739, Jul. 2008, https://doi.org/10.1016/j.ymssp.2008.06.009

[5] S. Yin, X. Li, H. Gao, and O. Kaynak, "Data-based techniques for fault diagnosis," *IEEE Transactions on Industrial Electronics*, Vol. 61, No. 11, pp. 6406–6418, 2014.

[6] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, Vol. 521, No. 7553, pp. 436–444, May 2015, https://doi.org/10.1038/nature14539

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, Vol. 60, No. 6, pp. 84–90, May 2017, https://doi.org/10.1145/3065386

[8] R. Salmorbekova and M. Talipov, "Digital risk matrix and safety management workflow for airport infrastructure in developing countries: a data-driven prioritization approach," *Vibroengineering Procedia*, Vol. 62, pp. 653–661, Jun. 2026, https://doi.org/10.21595/vp.2026.26121

[9] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[10] C. Elkan, "The foundations of cost-sensitive learning," in *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 973–978, 2001.

[11] G. Qian and J. Liu, "A comparative study of deep learning-based fault diagnosis methods for rotating machines in nuclear power plants," *Annals of Nuclear Energy*, Vol. 178, p. 109334, Jul. 2022, https://doi.org/10.1016/j.anucene.2022.109334

[12] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mechanical Systems and Signal Processing*, Vol. 100, pp. 439–453, Aug. 2017, https://doi.org/10.1016/j.ymssp.2017.06.022

[13] Y. Wang, D. Li, L. Li, R. Sun, and S. Wang, "A novel deep learning framework for rolling bearing fault diagnosis enhancement using VAE-augmented CNN model," *Heliyon*, Vol. 10, No. 15, p. e35407, Jul. 2024, https://doi.org/10.1016/j.heliyon.2024.e35407

[14] N. Rezazadeh, F. Caputo, A. Aversano, G. Lamanna, A. de Luca, and D. Perfetto, "Prototype-attention domain adaptation for explainable bearing fault diagnosis," *Procedia Structural Integrity*, Vol. 80, pp. 411–417, 2026, https://doi.org/10.1016/j.prostr.2026.02.039

[15] D. J. Hand, "Measuring classifier performance: a coherent alternative to the area under the ROC curve," *Machine Learning*, Vol. 77, No. 1, pp. 103–123, 2009, https://doi.org/10.1007/s10994-009-5119-5

[16] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: a review and roadmap," *Mechanical Systems and Signal Processing*, Vol. 138, p. 106587, Jan. 2020, https://doi.org/10.1016/j.ymssp.2019.106587

[17] J. Quiñonero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, "Dataset shift in machine learning," *Journal of the Royal Statistical Society Series A: Statistics in Society*, Vol. 173, No. 1, pp. 274–274, 2009, https://doi.org/10.1111/j.1467-985x.2009.00624_10.x

[18] M. Sugiyama, M. Krauledat, and K. R. Muller, "Covariate shift adaptation by importance weighted cross validation," *Journal of Machine Learning Research*, Vol. 8, pp. 985–1005, 2007, https://doi.org/10.1007/11861898_36

[19] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Pearson, 2014.

[20] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 2009.

[21] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[22] J. V. Carter, J. Pan, S. N. Rai, and S. Galandiuk, "ROC-ing along: evaluation and interpretation of receiver operating characteristic curves," *Surgery*, Vol. 159, No. 6, pp. 1638–1645, Mar. 2016, https://doi.org/10.1016/j.surg.2015.12.029

[23] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, Vol. 27, No. 8, pp. 861–874, Dec. 2005, https://doi.org/10.1016/j.patrec.2005.10.010

[24] B. Zadrozny and C. Elkan, "Transforming classifier scores into accurate multiclass probability estimates," in *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 694–699, 2002, https://doi.org/10.1145/775047.775151

[25] A. Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in *Proceedings of the 22nd International Conference on Machine Learning – ICML '05*, pp. 625–632, 2006, https://doi.org/10.1145/1102351.1102430

[26] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing*, Vol. 21, No. 6, pp. 2560–2574, Jan. 2007, https://doi.org/10.1016/j.ymssp.2006.12.007

[27] Z. Chen, A. Mauricio, W. Li, and K. Gryllias, "A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks," *Mechanical Systems and Signal Processing*, Vol. 140, p. 106683, Feb. 2020, https://doi.org/10.1016/j.ymssp.2020.106683

[28] A. Mamatov, X. Sotvoldiyev, M. Talipov, S. Shaumarov, K. Gafarbayli, and D. Bekmirzaev, "Metrological calibration and uncertainty evaluation of MEMS accelerometers for structural health monitoring in seismic regions," *Vibroengineering Procedia*, Vol. 62, pp. 140–144, Jun. 2026, https://doi.org/10.21595/vp.2026.26360

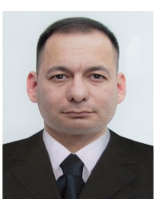
[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference for Learning Representations*, 2015, https://doi.org/10.48550/arxiv.1412.6980

[30] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proceedings of the 23rd International Conference on Machine Learning – ICML '06*, pp. 233–240, 2006, https://doi.org/10.1145/1143844.1143874

[31] I. Mirzaev, D. Bekmirzaev, E. Kosimov, E. An, N. Nishonov, and M. Talipov, "Seismodynamics of segmented underground pipeline systems based on real earthquake records," in *AIP Conference Proceedings*, Vol. 3447, No. 1, p. 060001, May 2026, https://doi.org/10.1063/12.0044018

[32] N. R. Yusupbekov, S. M. Gulyamov, and M. Y. Doshchanova, "Neural identification of a dynamic model of a technological process," in *2019 International Conference on Information Science and Communications Technologies (ICISCT)*, pp. 1–5, 2020, https://doi.org/10.1109/icisct47635.2019.9011912

Malika Doshanova successfully defended her dissertation in 2022 at the Scientific Council for the Award of Scientific Degrees DSc.13/30.12.2021.T.142.01 under the Scientific Research Institute for the Development of Digital Technologies and Artificial Intelligence. She received the degree of Doctor of Philosophy (Ph.D.) in Technical Sciences, specializing in 05.01.02 – systematic analysis, control and information processing, with the dissertation titled "Methods and Algorithms for Decision-Making in Technological Processes and Production Control Systems."



Miraziz Talipov received the Ph.D. degree in occupational safety from Tashkent State Transport University, Tashkent, Uzbekistan, in 2023. He currently works at the Department of Scientific Publications, Tashkent State Transport University. His research interests include occupational safety in railway transport, vibration protection of personnel and ergonomics of rolling-stock service equipment.

ISSN PRINT 2351-5279, ISSN ONLINE 2424-4627

Madina Shaazizova successfully defended her dissertation in 2024 at the Scientific Council for the Award of Scientific Degrees DSc.03/26.05.2022.T.10.05 under the Tashkent Institute of Irrigation and Agricultural Mechanization Engineers National Research University. She received the degree of Doctor of Philosophy (Ph.D.) in Technical Sciences, specializing in 05.01.03 – theoretical foundations of informatics, with the dissertation titled "Algorithms for Early Detection of Breast Cancer."

Nilufar Mirzayeva is a researcher in computer science and software engineering, currently pursuing Doctor of Philosophy (Ph.D.) in Technical Sciences in the specialty 05.01.11 – digital technologies and artificial intelligence at the Tashkent University of Information Technologies named after Muhammad al-Khwarizmi. Her research focuses on intelligent analysis of video data using artificial intelligence. Her interests include machine learning techniques and the detection of signs of drowsiness and fatigue in humans using video data.

Shirin Karaxanova is a researcher in computer science and software engineering, currently pursuing a Doctor of Philosophy (Ph.D.) in Technical Sciences in the specialty 05.01.04 – mathematical and software support of computers, complexes and computer networks. She completed her doctoral coursework between 2022 and 2024 at Tashkent University of Information Technologies named after Muhammad al-Khwarizmi. Her research focuses on natural language processing (NLP), sentiment analysis, and the dynamic modeling of mass discontent based on textual data. Her interests include the development of intelligent systems for tracking temporal changes in public opinion using machine learning techniques and large-scale datasets.

Indira Khodzamuratova successfully defended her dissertation in 2023 at the Scientific Council for awarding academic degrees under the Mathematics Institute of the Academy of Sciences of the Republic of Uzbekistan (DSc.02/30.12.2019.FM.01.01). She was awarded the degree of Doctor of Philosophy (PhD) in physical and mathematical sciences in the specialty 01.01.06 – algebra, and defended her dissertation entitled "Effective Topological Spaces over Algorithmic Representations of Universal Algebras".