INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2025, VOL. 71, NO. 3, PP. 1-6
Manuscript received June 25, 2025; revised July 2025. doi: 10.24425/ijet.2025.153635

Development of fault detection system in irrigation pumping systems using machine learning methods with consideration of energy and water consumption

Gulnar Zholdangarova, and Waldemar Wójcik

***Abstract*—Pumping systems play an important role in agriculture because they provide the necessary level of irrigation needed to increase crop yields. Pump malfunctions result in equipment downtime, reduced efficiency of agricultural production and significant financial losses. Thus, the development of an early fault detection and diagnosis system leveraging sensor analytic, filtering techniques, and machine learning (ML) technologies constitutes a critical applied research challenge. The aim of this research is to develop and validate early fault detection and classification methods for pumping systems using advanced machine learning algorithms and sensor data analysis.**

***Keywords*—vibration signal; time series; earing fault; particle swarm optimization; normalization**

I. Introduction

TRADITIONAL diagnostic methods for pumping systems are based on manual and regular inspections. Vibration analysis [1], temperature analysis [2], wear analysis, motor current signature analysis [3] and acoustic emission analysis [3] have been the basis of many diagnostic methods. Vibration analysis is considered the most effective of these methods because it can provide significant information about anomalies. Their main drawbacks include low accuracy, inability to detect hidden issues in a timely manner, and heavy reliance on manual labor. By contrast, modern diagnostic methods provide adaptive filtering, feature extraction, and model training on large amounts of sensory data. Modern machine learning methods are automated and accurate, but require large volumes of labeled data and complex infrastructure for processing

Faults in irrigation systems are mainly detected through periodic manual inspections or basic monitoring based on threshold values. Predicting and monitoring maintenance in real time has been made possible by recent developments in machine learning and sensor technologies [5]. The introduction of the Internet of Things (IoT) enhances remote monitoring capabilities by facilitating the real-time transfer of sensor data to centralized systems. Sensor fusion techniques that combine data from multiple sensors will improve diagnostic accuracy. For example, high diagnostic accuracy can be achieved by fusing data from vibration, pressure, and temperature sensors [6]. However, noise and data redundancy remain significant challenges in sensor deployment. The application of dimensional reduction techniques such as PCA is necessary due to the large amount of data from different sensors [7].

Irrigation pumping systems utilize several sensors to monitor operating parameters. These include temperature, pressure, vibration and flow sensors. For example, studies have shown that vibration data is critical for detecting mechanical faults, and they have shown that the accuracy of fault detection using neural networks is 93 percent [8]. To monitor environmental conditions that affect pump performance, temperature and humidity sensors are also needed [9].

Leaks, clogs, pump wear and vibration are the most common pump problems. These faults cause changes in pressure, water flow, vibration and temperature in pump parts.

In developing an effective condition monitoring technique, one of the important steps is to choose the right approach. A model-based approach based on physical and mathematical models of the equipment provides lower accuracy for complex installations, while a data-driven machine learning approach is more promising. Neural networks, RF, SVM, KNN and ensemble models are all machine learning algorithms with great potential [10, 11].

II. Experimental Setup

Having a suitable dataset is essential for building a classification algorithm. There are open datasets such as “NASA repository” and experimental platform “PROGNOSTIA”. However, in this study, experimentally obtained datasets were used. For this purpose, an experimental setup consisting of a three-phase induction motor with variable speed capability, a pumping equipment rotating element bearing, and a power transmission system was designed. The pumping equipment bearing was mounted in the system in such a way that it could be easily replaced. There were also opportunities to change the balancing mass of the system. The development of the experimental setup was one of the objectives of this research work. It was necessary to

This work was carried out at the expense of grant financing of scientific research for 2024-2026 under the project AP23490529 “Development of information system and mathematical models for monitoring and load forecasting of electric power systems on the basis of hybrid technologies”

Gulnar Zholdangarova is with L. N. Gumilev Eurasian National University, Astana, Kazakhstan (e-mail: zholdangarova9443@gmail.com).

Waldemar Wójcik is with Lublin University of Technology, Lublin, Poland (e-mail: waldemar.wojcik@pollub.pl)

ensure the reliability of the various components used to develop this setup. The effectiveness of the machine learning approach is highly dependent on the quality of the data.

Figure 1 shows the experimental setup. A three phase induction motor with a rated power of 0.5 hp and speed up to 6000 rpm was used as the power source in the experimental setup. A frequency converter (FC) with rated frequency from 0 to 599 Hz was installed in front of the induction motor to operate the machine under different operating conditions.

The power supply from the primary source was supplied to the motor through this IF. The shaft was connected to the motor using a flexible joint. Four pumping equipment bearings carried the static and dynamic load on the shaft. Flexible coupling and mounting systems were used for easy replacement of the pumping equipment bearings and the shaft. In the experiment, it was possible to create faults in all bearings of pumping equipment, this paper considers one bearing of pumping equipment to establish a reference point. A vibration analyzer with a measurement range of ±5 VAC, a frequency range of 2 to 10 kHz and a maximum sampling rate of 25.6 kHz based on the Raspberry Pi 4 platform was used. Acceleration data was measured using a Ronds accelerometer sensor with a measurement range of ±80 g, a frequency range of 0.7 to 10,000 Hz and a resonance frequency of about 30 kHz.

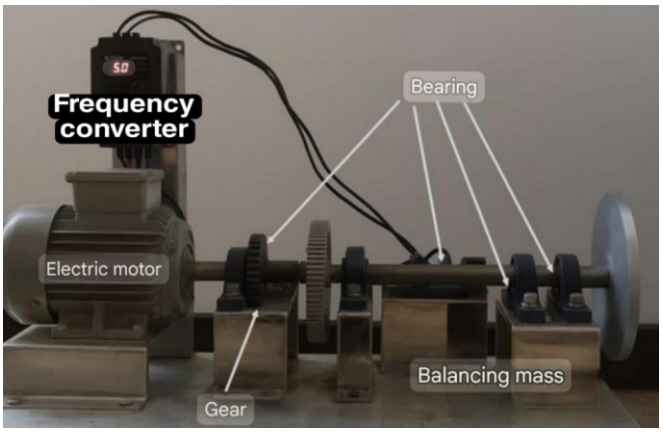

Fig.2.Experimental setup diagram with motor and converter

III. Methodology of work

In the experimental part, a modular hardware-software system based on Raspberry Pi 4 was developed to monitor irrigation pump parameters using vibration, water flow, and electrical sensors. Real-time sensor data were preprocessed on the Raspberry Pi and transferred to a PC, where filtering, normalization, and feature extraction were performed. Key features were calculated in time and frequency domains using statistical analysis and FFT to support machine learning–based fault diagnosis. The overall methodology of the work can be divided into several stages, as illustrated in Figure 2. All stages are described below.

The obtained attributes served as input data for machine learning models implemented at the stages of diagnostics and forecasting. As a result, the system provided not only the classification of technical condition of pumps (normal / faulty), but also the calculation of energy and water consumption parameters, allowing to assess the efficiency of the equipment as a whole. Such integration of sensor monitoring with intelligent analytics creates prerequisites for building full-fledged predictive maintenance systems in the field of agricultural water technologies.

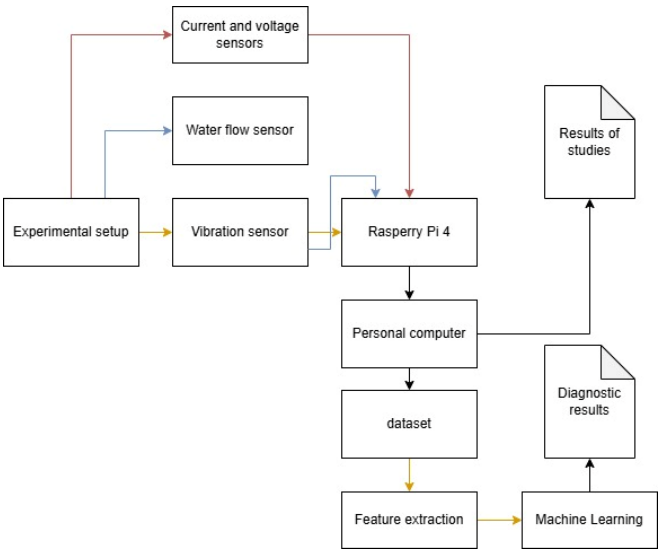

Fig.2. Structure of the monitoring and diagnostics system

As mentioned above, only one pumping equipment bearing was considered as test pumping equipment. All data of the serviceable and faulty bearing unit were collected according to a predefined data generation plan as shown in Table I.

TABLE I
DATA GENERATION PLAN

Number	Pumping equipment bearing condition	Motor speed, Hz
1	Without defect	20
2	Faulty outer ring	10, 20, 30
3	Inner ring malfunction	10, 20, 30
4	Cage malfunction	10, 20, 30

For example, Outer_Race_10 (OR-10) denotes a set of outer ring fault data collected at a frequency of 10 Hz. The generated vibration data were collected using an accelerometer mounted on the pump equipment casing.

The experimental data obtained from the rig using the accelerometer and transducers is saved locally on the Raspberry Pi 4 platform as time series (oscillograms). These time domain signals are then converted into .txt format and datasets (dataset) in .csv format. In the next step, various features are extracted from the original time domain signals such as:

- Kurtosis (measure of tailedness or peakedness),
- Amplitude factor (crest factor),
- Form factors (form factors),
- and others – using appropriate mathematical formulas [12].

Based on the extracted features, a feature matrix is formed.

The subsequent stage is to develop models (machine learning algorithm) for fault detection. In the last stage, the model was used for classification and its performance was evaluated using the following metrics:

- precision (accuracy),
- precision,
- recall (completeness),
- F1 score.

IV. MACHINE LEARNING METHOD FOR FAULT DETECTION

As part of this work, a fault detection approach was implemented based on the training of a Support Vector Machine (SVM) model with automatic hyperparameter optimization using the Particle Swarm Optimization (PSO) algorithm. The multi-class SVM, a classification algorithm based on the principle of structural risk minimization (SRM) [13], is employed in this study. SVM is efficient for small samples and searches for an optimal separating hyperplane that maximizes the gap between classes. The data points closest to the hyperplane are known as support vectors. For any given data set (u_1, v_1), (u_2, v_2),..., (u_n, v_n), $u_i \in R^n$ are considered as inputs, where u and v represent the corresponding data points of the reference vectors, a $v_i \in (-1,+1)$ are the outputs for each u_i.

A hyperplane in feature space is defined as:

$$f(x) = \omega^T s + b = 0, \quad (1)$$

where $\omega \in R^d$ is the normal to the hyperplane and $b \in R$ denotes the displacement (bias).

For linearly separable data:

$$y_i(\omega^T x_i + b) \geq 1, \forall i. \quad (2)$$

The optimizing problem is to maximize the gap between classes:

$$\text{mimimize} \frac{1}{2}\|\omega\|^2, \quad (3)$$

under the conditions

$$y_i(\omega^T x_i + b) \geq 1. \quad (4)$$

Variable deviations are introduced to allow for errors (noise in the data) $\xi i \geq 0$. Then the optimization problem:

$$\text{mimimize} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n} \xi i, \quad (5)$$

where:

$$y_i(\omega^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \quad (6)$$

and C>0 denotes the regularization parameter that controls the penalty for misclassification.

Particle Swarm Optimization (PSO) algorithm belongs to stochastic and population-based optimization methods. Instead of a single solution, it works with multiple particles (particles), each representing a possible solution to the problem. [14].

Each particle has:

- position $x_i \in R_n$ – the current solution;
- velocity $v_i \in R_n$ – the direction and magnitude of movement in the solution space;
- personal best solution p_i – the best position found by it;
- global best solution g – the best solution among all particles.

The velocity of each particle at iteration t is updated according to the following formula:

$$v_i(t+1) = \omega \cdot v_i(t) + c_1 \cdot r_1 \cdot \big(p_i - x_i(t)\big) + c_2 \cdot r_2 \cdot \big(g - c_1 \cdot x_i(t)\big), \quad (7)$$

The position is updated as:

$$x_i(t+1) = x_i(t) + v_i(t+1), \quad (8)$$

where: ω is the inertial coefficient (aspiration to the current direction), c_1, c_2 are cognitive and social coefficients, $r_1, r_2 \in [0,1]$ are some random values (ensure stochasticity), p_i is the personal best position (exploration), and g denotes the social experience (exploitation).

The proposed PSO-SVM approach reduces the computational complexity compared to the brute force method, has the ability to avoid local minima and demonstrates improved classification accuracy, which makes it a preferred tool in the tasks of intelligent diagnostics of technical systems.

To objectively evaluate the effectiveness of the developed classification model, we used standard metrics based on the error matrix (confusion matrix). The error matrix is a table of size $K \times K$ (where K is the number of classes), each cell of which M_{ij} shows the number of objects belonging to true class i, but assigned to class j by the model. The following metrics are calculated on the basis of the error matrix:

- Accuracy (9) reflects the proportion of correctly classified objects to all objects.
- Recall (10), or sensitivity, measures the ability of the model to correctly identify objects of a particular class, A high Recall value is important in tasks where a missed error (false negative) is critical (e.g., fault diagnosis).
- Precision (11) shows how many of the objects, actually belong to this class. A high Precision value is critical when false positives are undesirable (e.g., false alarms).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP+FN}. \quad (10)$$

$$\text{Precision} = \frac{TP}{TP+FP}. \quad (11)$$

F1-score (12) is a harmonic mean between Precision and Completeness, providing a balanced score. F1-score is particularly useful in unbalanced classes, where high accuracy on a single metric does not guarantee overall classification performance.

$$\text{F1score} = \frac{\text{Precision} \times \text{Recal}}{\text{Precision} + \text{Recal}}. \quad (12)$$

Thus, a comprehensive analysis of the diagnosis model was performed using Accuracy, Precision, Recall and F1-score metrics, which provided a comprehensive evaluation of both generalizability and robustness of classification on all types of pumping equipment faults. This is especially important in technical diagnostic tasks where different types of faults have different criticality. The results of the PSO-SVM model classification quality evaluation on the metrics of precision, recall and F1-score on the test sample are presented in Table II.

High F1-score values (>0.90) for all classes confirm the effectiveness of the model in bearing fault differentiation tasks. It is particularly noteworthy that perfect accuracy and completeness (1.00) was achieved for the Inner Race Fault class, indicating stable identification of this type of fault.

TABLE II
CLASSIFICATION METRICS FOR EACH CLASS

Class	Precision	Recall	F1-Score	Support
Normal	0.97	1.00	0.98	15
Outer Race Fault	0.95	0.92	0.93	13
Inner Race Fault	1.00	1.00	1.00	12
Cage Fault	0.89	0.85	0.87	20
Macro average	0.95	0.94	0.94	—
Weighted average	0.95	0.95	0.95	60
Normal	0.97	1.00	0.98	15
Outer Race Fault	0.95	0.92	0.93	13

V. RESEARCH RESULTS

Examples of experimental data are shown in the following figures. Figure 3 shows the time diagram of the intact bearing – Sensor 1, and Figure 7 shows the time diagram of the Faulty bearing – Sensor 2. For the intact bearing the maximum amplitude was 0.0035 m/s^2 while for the faulty bearing the amplitude increased to 0.006 m/s^2

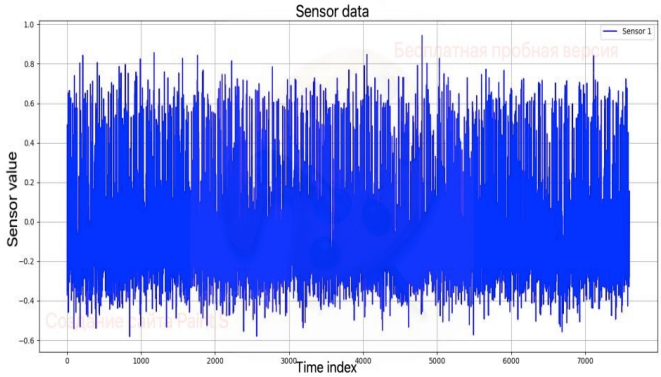

Fig.3. Filtered vibration data from Sensor 1 (serviceable)

In Figure 3 on the abscissa axis the time index is plotted (in arbitrary units), corresponding to discrete samples of the signal, and on the ordinate axis - normalized values of vibration amplitude. The signal does not contain pronounced impulse emissions characteristic of contact surface damage, and the vibrations remain within the established amplitude range. The vibration amplitude is within the range from –0.5 to +0.9, and the signal structure demonstrates statistical stability, which indicates the absence of signs of mechanical malfunctions

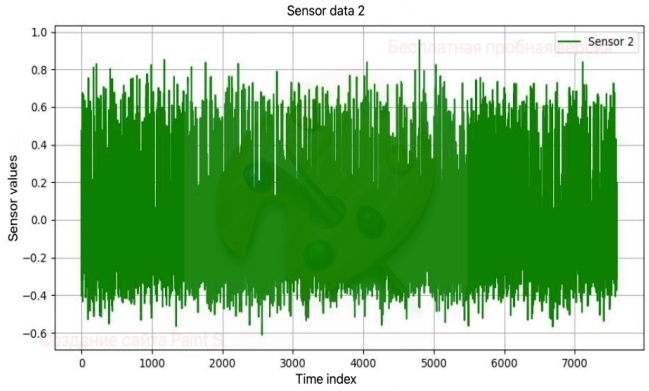

Fig.4. Vibration sensor data from Sensor 2 (faulty)

The graph in Figure 7 is constructed by analogy with Figure 3: the time index is plotted along the abscissa axis, and the normalized vibration amplitudes are plotted along the ordinate axis. In contrast to the signal of serviceable condition, this graph is characterized by higher irregularity and the presence of local anomalous emissions. Denser impulse oscillations are especially clearly observed, which indicates the presence of shock interactions between rolling elements and damaged elements of bearing shells.

The figure shows the stage of preliminary processing of the vibration signal received from Sensor 1 using the filtering steps. The blue color represents the original signal, and the red line represents the result of filtering performed using a band-pass filter adjusted to remove high-frequency noise and possible measurement artifacts. The application of filtering at this stage allowed to significantly improve the signal-to-noise ratio and increase the reliability of the subsequent extraction of features used for machine learning.

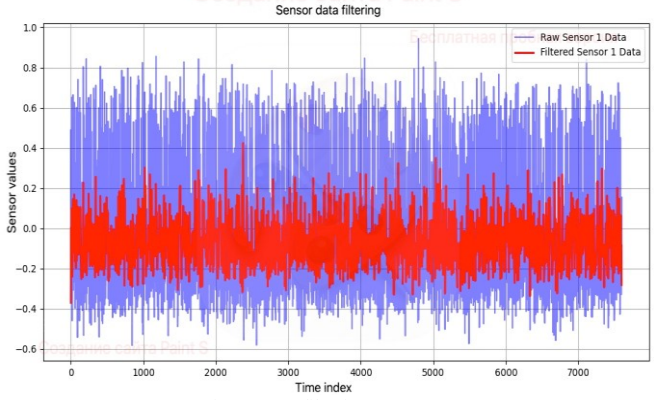

Fig.5. Data filtering Sensor 1

Figure 6 shows the result of filtering the vibration signal from Sensor 2 that detected a bearing condition with a fault. The green area represents the original data and the yellow-orange graph of the filtered signal. Filtered using a band-pass filter designed to suppress high frequency noise and stabilize the signal. After filtering, it was possible to significantly suppress spurious vibrations while retaining informative components related to the fault dynamics. Filtering the vibration signal improved data quality and provided a more accurate differentiation between normal and faulty modes of operation of the mechanism.

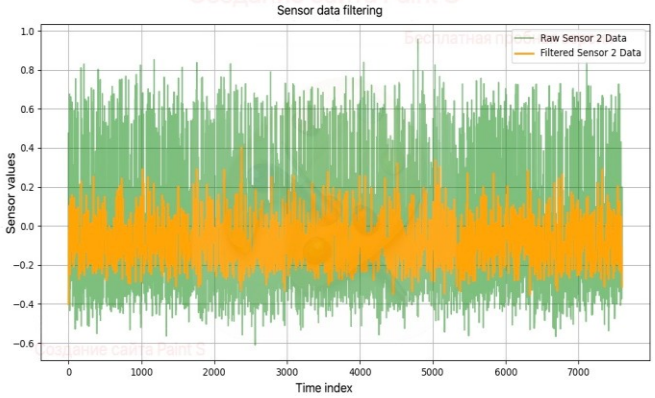

Fig.6. Data filtering Sensor 2

The process of normalization of vibration signals obtained from Sensor 1 and Sensor 2 is shown in Figure 7. As can be seen from the graphs, the standardized signals retain the shape and dynamics of the original vibrations, but are scaled in such a way that the vibrations fit predominantly in the range from -2 to +3.

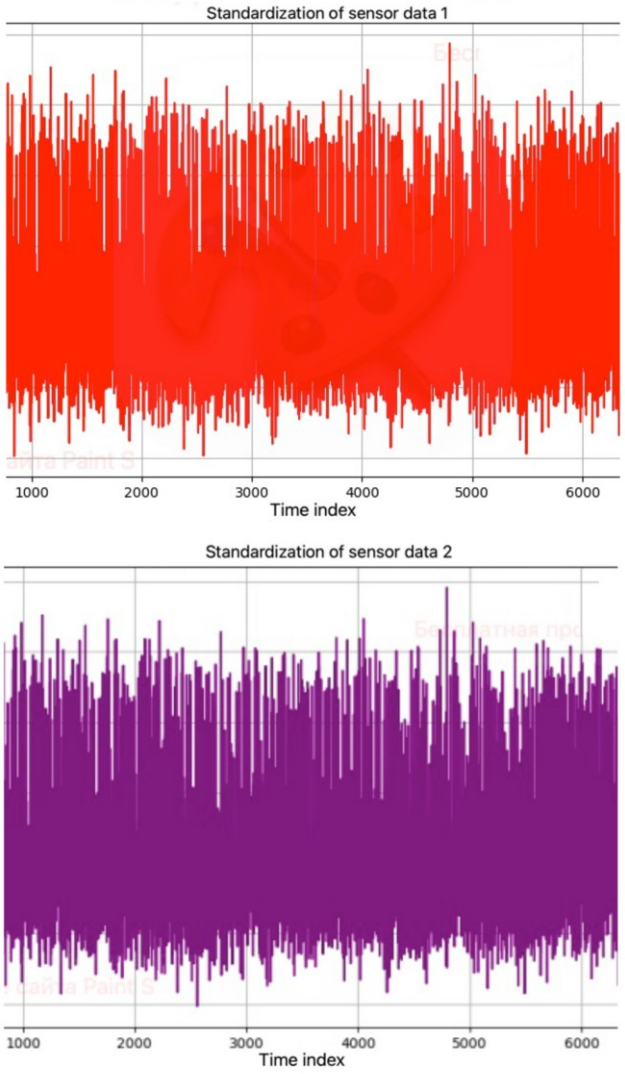

Fig.7. Normalization of Sensor 1 and Sensor 2 data

To analyze the characteristics of vibration signals recorded from rotating equipment for the purpose of diagnosing the technical condition of bearings, a decomposition of the time series was performed (Figure 8). This approach allows identification of the structural components of the signal and better understand the nature of emerging deviations associated with potential faults. Data preprocessing includes filtering, handling missing values, feature extraction, signal processing, data normalization, data scaling, and selection of the optimal feature subset. Such decomposition functions as an adaptive filtering technique, enabling the isolation of the most informative signal components for further analysis. This enhances model resilience to noise and improves sensitivity to early indicators of system faults.

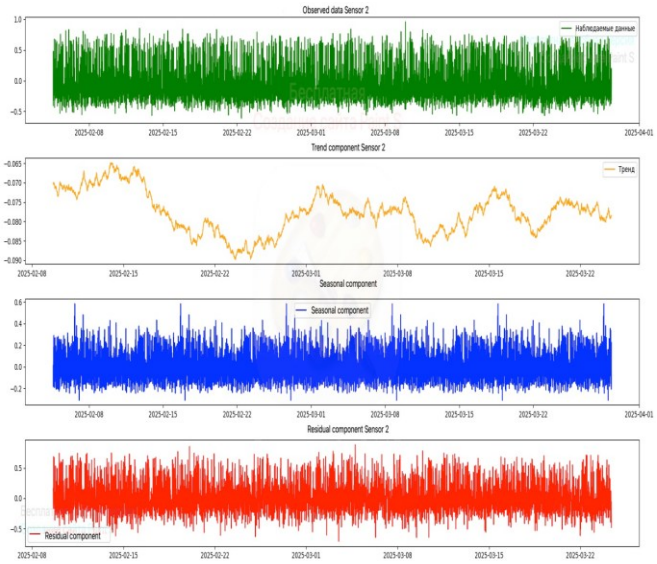

Fig. 8. Normalization of Sensor 1 and Sensor 2 data.

In particular, we considered a time series of vibration data obtained from a sensor installed near a bearing suspected to have mechanical damage (raceway wear). The time series plot includes the following components:

Initial time series (Observed) - represents a sequence of recorded vibration data in a selected projection (usually along the X, Y or Z axes). For faulty bearings, such data may contain pulse bursts, increased amplitude and characteristic frequencies associated with the type of damage (e.g. BPFO, BPFI, BSF, FTF frequencies).

Trend component (Trend) - reflects long-term changes in the overall vibration level that may be associated with progressive bearing wear, misalignment, center of mass displacement, or accumulation of mechanical faults. In the context of a defective bearing, the observed trend may indicate a gradual increase in vibration activity as the condition of the unit deteriorates.

Seasonal component (Seasonal) - displays periodic oscillations associated with shaft rotation and regular passage of defective elements through the contact zone. For a damaged bearing, the seasonal component may include characteristic repetitive pulses corresponding to the frequencies of fault occurrence on the inner and outer rings. The frequency of these oscillations is often related to the calculated diagnostic frequencies of the bearing and can serve as a key sign of a fault.

Residual component (Residual) – includes high-frequency noise and irregular fluctuations not explained by trend or seasonality. In the case of bearing faults, the residuals may contain sporadic shock signals due to initial stages of faults or unstable operation. Increased dispersion and amplitude of this component may be indicators of deterioration.

Thus, the decomposition of time series within the framework of the tasks of diagnostics of vibration state of bearings allows:

- localize typical characteristic fault frequencies (by seasonal component),
- track fault progression (by trend),
- analyze unstable noise characteristics (by residuals), which together contributes to accurate and timely detection of failures.

The classification results are presented in Table III. The obtained results can be used to evaluate the prediction accuracy of the classification algorithm. The accuracy value shows the ratio of the actual predicted true value to the total predicted true value.

TABLE III
CLASSIFICATION METRICS BY CLASS (PSO-SVM)

Class	Precision	Recall	F1-Score
Cage_Fault_10	0.9359	0.9733	0.9542
Cage_Fault_20	0.8955	0.8000	0.8451
Cage_Fault_30	0.8462	0.8800	0.8627
Inner_Race_10	1.0000	1.0000	1.0000
Inner_Race_20	1.0000	1.0000	1.0000
Inner_Race_30	0.9600	0.9600	0.9600
Normal_20	0.9740	1.0000	0.9868
Outer_Race_10	1.0000	0.9867	0.9933
Outer_Race_20	0.8267	0.8267	0.8267
Outer_Race_30	0.9474	0.9600	0.9536
Accuracy	-	-	**0.9387**
Macro Average	0.9386	0.9387	0.9383
Weighted Average	0.9386	0.9387	0.9383

The classification results of the PSO-SVM model into ten subtypes of bearing states. The model showed high values of metrics in all classes, including the critical states Inner_Race_10 and Outer_Race_10, for which the F1-score exceeds 0.99. The mean values (macro and weighted) of F1-score exceed 0.93, which confirms the high stability of the model and the balanced classification.

VI. CONCLUSION

In this study, a diagnostic framework was developed that integrates machine learning algorithms, signal filtering techniques, and optimization methods to enable the early identification of faults in pump systems. For this purpose, an experimental setup was designed and fabricated, taking into account all design parameters, in order to obtain data as close as possible to real industrial applications. The experimental data obtained was analyzed to separate serviceable and damaged bearings depending on the severity of the fault.

During the experiments, the data were recorded under acoustically isolated conditions to eliminate the influence of external noise, and the experimental setup was securely mounted to minimize its own vibrations. The time domain signals were saved as images, which were subsequently converted into numerical format using the in-house software of the vibration analyzer. The data then underwent signal processing steps including filtering, interpolation of missing values, scaling and normalization. Statistical features were extracted from the signals, after which the most informative ones were selected on the basis of the correlation matrix. Correlation analysis showed a complete linear relationship between the root mean square (RMS) and standard deviation (correlation coefficient of 1.0), and a strong positive correlation between the crest factor and the skewness coefficient ($\rho = 0.83$).

The generated feature matrix was used in the PSO-SVM model. The classification accuracy achieved using the support vector algorithm with parameter optimization by particle swarm method was 93.9%, which is almost 2% higher than the accuracy of SVM tuned using the classical grid search method with cross-validation. The improved quality of the proposed hybrid model is explained by its high robustness to noise in the training data and lower tendency to overtraining compared to traditional algorithms.

REFERENCES

[1] Y.Wang, J.Xiang, R.Markert, and M.Liang, “Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications,” Mech. Syst. Signal Process.vol. 66, pp. 679–698, Mar. 2016. https://doi.org/10.1016/j.ymssp.2015.04.039

[2] 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, pp.106587, 2020. http://dx.doi.org/10.1016/j.ymssp.2019.106587

[3] N. Mehala, “Condition Monitoring and Fault Diagnosis of Induction Motor Using Motor Current Signature Analysis,” Electr. Eng.,vol. 2, 175. 2010.

[4] K.Du, X.Li, M.Tao, and S.Wang, “Experimental study on acoustic emission (AE) characteristics and crack classification during rock fracture in several basic lab tests,” International Journal of Rock Mechanics and Mining Sciences, vol. 133, pp.104411, 2020. https://doi.org/10.1016/j.ijrmms.2020.104411

[5] T. Chen, F. Dong, H. Wang, Z. Xu, and E. Liu, “Construction of Intelligent Fault Diagnosis System for Pumping Station Based on Fusion Algorithm,” 454–460. (2024).

[6] G.Chen, L.Wang, H.Yang, P.Wang, J.Wei, and J.Bao, “Health assessment of water pumps using high-dimensional monitoring data,” Water Science & Technology Water Supply, vol. 23(2), 2023. http://dx.doi.org/10.2166/ws.2023.244

[7] D. Loukatos, M. Kondoyanni, G. Alexopoulos, C. Maraveas, and K. G. Arvanitis, “On-Device Intelligence for Malfunction Detection of Water Pump Equipment in Agricultural Premises: Feasibility and Experimentation,” Sensors, vol. 23(2), 839, 2023. https://doi.org/10.3390/s23020839

[8] S. Liang, P. Liu, S. Zhang, and Z. Wu, “Research on Fault Diagnosis of Agricultural IoT Sensors Based on Improved Dung Beetle Optimization–Support Vector Machine,” Sustainability, vol. 16(22), pp. 10001, 2024. https://doi.org/10.3390/su162210001

[9] S. Sahoo, A. Singh, and M. K. N. Kumari, “Identifying Anomalies in Water Pump Systems Using Machine Learning and an Integrated Ensemble Method,” in proc. ICDSIS, pp. 1-6, 17-18 May 2024. http://dx.doi.org/10.1109/ICDSIS61070.2024.10594390

[10] M. A. Febriantono, N. W. Prasetya, S. Sidharta, “Intelligent irrigation management system based on iot and machine learning,” AIP Conf. Proc., vol. 2927, pp. 040002, 2024. https://doi.org/10.1063/5.0193464

[11] D.K. Saha, S. Ahmed, and S. Shaurov, “Different Machine Maintenance Techniques of Rotary Machine and Their Future Scopes: A Review,” In Proceedings of the 2019 4th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, pp. 1–6, 20–22 December 2019.

[12] Q.Jiang,; F.Chang, “A novel rolling-element bearing faults classification method combines lower-order moment spectra and support vector machine,” Journal of Mechanical Science and Technology, vol. 33, pp. 1535–1543, 2019. https://doi.org/10.1007/s12206-019-0305-2

[13] R. Poli, J. Kennedy, and T. Blackwell, “Particle swarm optimization,” Swarm Intell, vol. 1, pp. 33–57, 2007. https://doi.org/10.1007/s11721-007-0002-0 2007