Вестник РУДН. Серия: Инженерные исследования
RUDN Journal of Engineering Research

2025;26(3):273–287

ISSN 2312-8143 (Print); ISSN 2312-8151 (Online)
journals.rudn.ru/engineering-researches

DOI: 10.22363/2312-8143-2025-26-3-273-287
EDN: YOXOFH

Research article / Научная статья

Comparative Performance of Machine Learning Classifiers in Detecting Vibration Anomalies in Industrial Power Systems

Al-Tekreeti Watban Khalid Fahmi, Kazem Reza Kashyzadeh✉, Siamak Ghorbani, Siamak Ghorbani, Sergei A. Kupreev, Oleg E. Samusenko

RUDN University, *Moscow, Russian Federation*
✉ reza-kashi-zade-ka@rudn.ru

Article history
Received: March 13, 2025
Revised: April 30, 2025
Accepted: May 15, 2025

Conflicts of interest
The authors declare that there is no conflict of interest.

Data Availability Statement:
The datasets can be made available upon request from the authors.

Abstract. This study examines methodologies for detecting abnormalities in Combined Cycle Power Plants (CCPPs) through application of vibration signal analysis and machine learning algorithms. Models' performances were evaluated using different key metrics. The results indicated that the Random Forest classifier, particularly in combination with ECPT data, exhibited superior performance, achieving perfect scores across all metrics. It highlights the robustness of the Random Forest algorithm when applied to ECPT data, making it the most effective approach for vibration anomaly detection. The K-NN classifier demonstrated satisfactory performance when applied to AS and BTT data, attaining accuracy scores of 0.49 and 0.52, respectively; however, it exhibited limitations in handling diverse data distributions, as reflected in its lower accuracy of 0.44 with LDV data. Both GBM and SVM performed suboptimal, with GBM achieving a maximum accuracy of 0.52 with AS data, while SVM attained the highest accuracy of 0.49 with the same technique. Findings underscore the critical importance of selecting an appropriate combination of machine learning models and vibration measurement techniques to enhance the accuracy of anomaly detection. Eventually, the Random Forest algorithm is well suited for complex datasets with varied patterns, while K-NN may serve as an efficient alternative for simpler, more uniform data.

Keywords: Vibration data, Fault diagnosis, Machine learning classification, Condition monitoring, Combined cycle power plants, CCPP, Predictive maintenance

Authors' contribution

Fahmi A.T.W.K. — methodology, software, validation, writing-original draft preparation; *Reza Kashyzadeh K.* — methodology, formal analysis, investigation, writing-review and editing, supervision, project administration; *Ghorbani S.* — methodology, investigation, writing-original draft preparation; *Kupreev S.A., Samusenko O.E.* — writing and editing. All authors read and approved the final version of the article.

For citation

Fahmi ATWK, Reza Kashyzadeh K, Ghorbani S, Kupreev SA, Samusenko OE. Comparative performance of machine learning classifiers in detecting vibration anomalies in industrial power systems. *RUDN Journal of Engineering Research.* 2025; 26(3):273–287. http://doi.org/10.22363/2312-8143-2025-26-3-273-287

Сравнительная эффективность классификаторов машинного обучения при обнаружении аномалий вибрации в промышленных энергосистемах

А.Т.В.Х. Фахми, К. Реза Каши Заде✉, С. Горбани, С.А. Купреев, О.Е. Самусенко

Российский университет дружбы народов, *Москва, Российская Федерация*
✉ reza-kashi-zade-ka@rudn.ru

История статьи

Поступила в редакцию: 13 марта 2025 г.
Доработана: 30 апреля 2025 г.
Принята к публикации: 15 мая 2025 г.

Заявление о конфликте интересов

Авторы заявляют об отсутствии конфликта интересов.

Заявление о доступности данных:

Данные могут быть предоставлены по запросу авторов.

Аннотация. Изучены методологии обнаружения отклонений в электростанциях комбинированного цикла посредством применения анализа сигналов вибрации и алгоритмов машинного обучения. Результаты показали, что метод случайного леса, особенно в сочетании с данными вихретоковых датчиков приближения, продемонстрировал превосходную эффективность, достигнув идеальных результатов по всем показателям. Это подчеркивает надежность алгоритма случайного леса при применении к данным вихретоковых датчиков приближения, что делает его наиболее эффективным подходом для обнаружения аномалий вибрации. Классификатор K-NN продемонстрировал удовлетворительную эффективность при применении к данным датчиков ускорения и датчики синхронизации кромки лопатки, достигнув показателей точности 0,49 и 0,52 соответственно; однако он продемонстрировал ограничения при обработке различных распределений данных, что отражено в его более низкой точности 0,44 с данными лазерных доплеровских виброметров. Машина для повышения градиента и метод опорных векторов показали неоптимальные результаты, причем машина для повышения градиента достигла максимальной точности 0,52 с данными датчиков ускорения, в то время как метод опорных векторов достиг наивысшей точности 0,49 с той же методикой. Результаты подчеркивают критическую важность выбора подходящей комбинации моделей машинного обучения и методов измерения вибрации для повышения точности обнаружения аномалий. В итоге алгоритм случайного леса хорошо подходит для сложных наборов данных с разнообразными моделями, в то время как K-NN может служить эффективной альтернативой для более простых и однородных данных.

Ключевые слова: данные о вибрации, диагностика неисправностей, классификация машинного обучения, мониторинг состояния, электростанции комбинированного цикла, ССРР, прогностическое обслуживание

Вклад авторов

Фахми А.Т.В.Х. — методология, программное обеспечение, проверка, написание — подготовка первоначального проекта; *Реза Каши Заде К.* — методология, формальный анализ, проведение исследования, написание — рецензирование и редактирование, руководство, администрирование проекта; *Горбани С.* — методология, проведение исследования, написание и подготовка первоначального черновика; *Купреев С.А., Самусенко О.Е.* — написание и редактирование текста. Все авторы ознакомлены с окончательной версией статьи и одобрили ее.

Для цитирования

Fahmi A.T.W.K., Reza Kashyzadeh K., Ghorbani S., Kupreev S.A., Samusenko O.E. Comparative performance of machine learning classifiers in detecting vibration anomalies in industrial power systems // Вестник Российского университета дружбы народов. Серия: Инженерные исследования. 2025. Т. 26. № 3. С. 273–287. http://doi.org/10.22363/2312-8143-2025-26-3-273-287

Introduction

Vibration analysis is a crucial aspect of condition monitoring in industries that rely on rotating equipment, such as petrochemical plants and power plants. In CCPPs, which utilize both gas and steam turbines, continuous monitoring and evaluation of vibration signals are essential to ensuring reliability and efficiency in power generation. Vibration analysis serves as an effective method for detecting early signs of mechanical failures, such as rotary imbalance, coupling misalignment, and component wear, before they lead to costly downtime, reduced efficiency, or catastrophic equipment failure [1; 2]. In a CCPPs, vibration monitoring is particularly critical for primary energy-generating machinery, such as gas turbines, where even minor faults like speed fluctuations, excessive vibration, or timing irregularities, can result in significant efficiency losses, increased fuel consumption, and unplanned shutdowns, ultimately affecting overall plant performance [3; 4]. Figure 1 presents various damages that occurred in the gas turbine of the Kirkuk power plant located in Iraq.

Multiple factors contribute to vibrations in gas turbines and other rotating equipment. Common issues include shaft unbalance, critical speed occurrence, rubbing, and shorted turns off. Each of these problems can be detected using specialized vibration analysis techniques [5; 6]. For example, shaft unbalance, a leading cause of high-amplitude vibrations, adversely affects bearings, shafts, and other rotating components, leading to increased maintenance costs and reduced operational efficiency. Figure 2 shows bearing damage due to misalignment in a gas turbine of the Kirkuk power plant located in Iraq.

a *b* *c*

Figure 1. Gas turbine damages in the Kirkuk power plant due to the occurrence of vibrations caused by: *a* and *b* — steam flow fluctuations; *c* — rubbing

Source: by Al-T.W.K. Fahmi

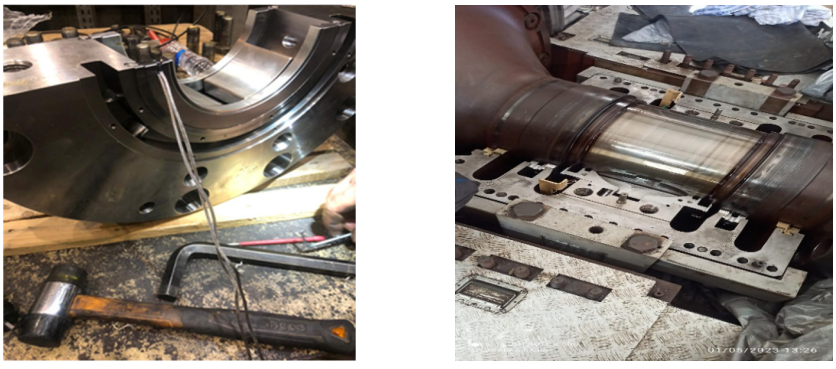

Figure 2. Bearing damage due to the misalignment in a gas turbine of the Kirkuk power plant located in Iraq

Source: by Al-T.W.K. Fahmi

Studies suggest that correcting unbalance and misalignment issues can greatly reduce power consumption in industrial machinery [7]. Similarly, critical speed resonance occurs when a machine operates at or near its natural frequency, causing excessive wear and potential failure. Preventative maintenance techniques, such as short-time Fourier transform (STFT), are often employed to detect critical speed issues during machine start-up and shutdown phases [8].

Various condition monitoring techniques are used to acquire the dynamic signatures of these mechanical defects. These include Eddy Current Proximity Transducers (ECPT), Accelerometer Sensors (AS), Blade Tip Timing (BTT), Laser Doppler Vibrometers (LDV), and Strain Gauges (SG), each offering unique advantages and limitations. ECPT, for instance, provides highly accurate displacement measurements for high-speed equipment but requires time consuming calibration [9]. Accelerometers are versatile and capable of measuring a wide range of vibration frequencies, though they are susceptible to electromagnetic interference [10]. BTT is a non-intrusive technique that provides high-resolution data on blade vibrations, but it is limited to blade tip measurements [11]. LDV is a highly sensitive contactless measurement method capable of detecting minute oscillations, though it requires sophisticated and costly equipment [12]. Finally, SGs are effective in measuring strain in structural components but require precise calibration and are influenced by temperature variations [13].

Recent advancements in machine learning (ML) techniques have introduced powerful new approaches to vibration data analysis and faults in complex industrial systems diagnostics. ML models such as Random Forest (RF), Gradient Boosting Machine (GBM), Support Vector Machine (SVM), and K-Nearest Neighbors (K-NN) have been widely used for processing large datasets, detecting abnormal patterns, and classifying vibration signals [14]. RF, an ensemble-based decision tree classifier, is particularly effective in handling high-dimensional and noisy data, making it well-suited for ECPT and AS datasets. GBM, another ensemble method, sequentially improves predictions by minimizing errors and is particularly useful for structured datasets, such as those generated by accelerometers and strain gauges [15]. SVM, a strong binary classifier, excels at finding optimal hyperplanes for separating different vibration patterns, though its performance is highly dependent on data structure and dimensionality [16]. K-NN, a distance-based classifier, operates under the assumption that a data point's classification is determined by its nearest neighbors. Despite its simplicity, K-NN performs well when dealing with densely clustered vibration data, such as those obtained from high-frequency techniques [17]. Its straightforward implementation and low computational requirements make it particularly useful for real-time industrial applications where processing power is limited. Given its efficiency, K-NN serves as a useful benchmark for comparing more complex models in vibration classification.

This study aims to evaluate and compare the effectiveness of RF, GBM, SVM, and K-NN classifiers in analyzing synthetic vibration data generated from five monitoring techniques: ECPT, AS, BTT, LDV, and SG. The synthetic data were designed to simulate real-world conditions by varying key parameters such as vibration frequency, amplitude, and noise levels, allowing for comprehensive testing across multiple operational scenarios. To enhance model performance, preprocessing techniques such as data labeling (normal vs. abnormal) and outlier removal were applied. The performance of each classifier was assessed using standard evaluation metrics, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic (ROC AUC) curve.

This study is guided by the following research questions:

1. Which machine learning classifier demonstrates the highest accuracy in detecting vibration anomalies across different monitoring techniques?

2. How do variations in vibration measurement techniques impact classifier performance?

3. Can a specific combination of machine learning models and monitoring techniques optimize vibration anomaly detection in CCPPs?

By addressing these questions, this research seeks to provide valuable insights into the integration of machine learning with traditional vibration analysis techniques, ultimately contributing to more efficient and predictive maintenance strategies in industrial settings [18].

1. A brief Description of the Most Important Vibration Measurement Techniques in Industry

The following is a brief description of the mechanism of the techniques used in this research to measure vibration in a gas turbine:

1.1. Eddy Current Proximity Transducers (ECPT)

ECPTs are widely used in power plants, particularly for monitoring the movement of the rotating machinery by detecting changes in proximity to an electromagnetic field. These sensors are preferred in high-speed applications and testing environments due to their reliability.

Research indicates that ECPTs can measure even the smallest displacement changes, making them suitable for tracking critical components such as turbine shafts and bearings [12]. However, ECPTs have some limitations: they can only sense movement in one direction, their calibration process is highly sensitive, time-consuming, and requires specialized equipment [19].

1.2. Accelerometer Sensors (AS)

Accelerometers are commonly used for vibration monitoring due to their high versatility and responsiveness across a wide frequency range. They operate based on a mass-spring system, generating an electrical signal proportional to acceleration, which is then analyzed to assess vibration characteristics. These devices are widely applied in power plants, particularly for monitoring turbines and evaluating structural integrity[1]. However, accelerometers are susceptible to external vibrations and electromagnetic noise, which can reduce measurement accuracy and necessitate frequent recalibration[2].

1.3. Blade Tip Timing (BTT)

BTT is an intrusive method commonly used to detect turbine blade vibrations in combined cycle power plants (CCPPs). It employs optical or microwave sensors placed around the rotor to detect the timing of blade tip passages. Research has shown that BTT can identify both high- and low-frequency vibrations without requiring any modifications to the turbine [11]. For example, Zhang et al. developed a microwave-based BTT system in which a patch antenna probe transmits and receives microwave signals reflected from the turbine blades, providing highly accurate measurements of blade dynamics [17].

1.4. Laser Doppler Vibrometers (LDV)

Laser Doppler Vibrometers (LDVs) are a contactless and a highly accurate method for measuring small vibrations using laser beams. They operate by detecting variations in the frequency of laser light reflected from a vibrating surface, enabling real-time observation of rotating equipment. A key advantage of LDVs is their high sensitivity and accuracy, making them suitable for detecting even the smallest vibrations [20]. However, their high cost and the complex structures required for implementing control algorithms may limit their use in large-scale industrial applications. Recent studies have aimed to improve the applicability of LDVs by enhancing signal detection for both low-frequency and high-frequency vibrations [21].

1.5. Strain Gauges (SG)

Strain gauges record distortions caused by strain or vibration by measuring variations in the electrical resistance of a small metal strip bonded

[1] *Accelerometer specifications: deciphering an accelerometer's datasheet*. Available from: https://www.scribd.com/document/650160365/Accelerometer-Specifications-Deciphering-an-Accelerometer-s-Datasheet (accessed: 01.03.2025).

[2] *Most Common Myths about Accelerometers and Frequency Range.* Available from: https://adash.com/articles/myths-accelerometers-frequency-range/ (accessed: 01.03.2025).

to a structure. They are particularly useful for monitoring large frameworks and support structures in power plants, as well as detecting structural distortions over time [13]. However, strain gauges typically have a low measurement range and can be affectedby temperature changes. Therefore, these sensors require precise calibration for accurate strain measurements. Recent developments in strain gauge technology aim to improve temperature compensation and enhance accuracy in harsh environments, making them more reliable for structural health monitoring throughout a structure's lifecycle[3].

2. Machine Learning Algorithms in Vibration Analysis

Machine learning methodologies have significantly improved the diagnosis and prediction of vibrations in complex industrial systems. Many studies highlight the advantages of using machine learning for fault detection, particularly in CCPPs, where early signs of equipment degradation can greatly impact plant reliability. The following is a brief description of the machine learning algorithms used in this research for the purpose of vibration analysis:

2.1. Random Forest (RF)

Random Forest is a widely used ensemble learning model known for its stability and ability to handle large numbers of features in vibration signal classification. It constructs multiple decision trees during training and integrates their results to improve classification efficiency. Previous research has demonstrated that the Random Forest algorithm performs well in detecting abnormal patterns from ECPT and accelerometer sensor data due to its low susceptibility to overfitting and its strong interpretability in large datasets with many variables. For example, one study showed that an RF model trained with synthetic vibration data for ECPT achieved 100% accuracy in exact measurement [14].

2.2. Gradient Boosting Machine (GBM)

GBM is an ensemble learning method composed of sequentially assembled decision trees, focusing on error minimization at each stage. This makes it particularly effective for the discrete datasets commonly used in vibration analysis. As misclassified cases are iteratively added to improve the model, GBM enhances its ability to identify minute patterns, such as those seen in accelerometer and strain gauge data. Studies have shown that GBM efficiently uncovers relationships within data and improves outlier detection by refining weak learners at each iteration step. In one study, GBM achieved an accuracy of 0.52 on AS data, demonstrating its effectiveness in classifying structured sensor data.

2.3. Support Vector Machine (SVM)

SVM is a well-known classification algorithm that selects the optimal hyperplane to separate data points. It is particularly effective in cases where binary classification is essential. Research has shown that SVM performs well in detecting vibration abnormalities, especially when using accelerometer and blade tip timing (BTT) data [22]. However, its performance dependents on the dataset structure and its time complexity increases with large datasets, which can hinder real-time applications in certain CCPP scenarios [16].

2.4. K-Nearest Neighbors (K-NN)

K-NN is a simple, instance-based learning model that classifies data points based on their similarity to neighboring data. Its simplicity makes it an ideal choice in scenarios where computational resources are limited, but fast classification is required. Studies indicate that K-NN performs well in density-based functions, such as accelerometer analysis and blade tip timing data, particularly when dealing with closely grouped datasets [18]. In one study, K-NN achieved accuracy rates of

[3] *Characteristics of a Strain Gauge Sensor. Bestech Australia.* Available from: https://www.bestech.com.au/blogs/characteristics-of-a-strain-gauge-sensor/ (accessed: 01.03.2025).

0.49 and 0.52 for AS and BTT data accordingly, respectively, highlighting its effectiveness as a lightweight, distance-based classifier in specific vibration monitoring scenarios [19].

Previous research confirms that combining vibration measurement techniques with fault classification algorithms significantly enhances fault diagnosis in CCPPs. Highly sensitive methods, such as ECPT and accelerometers, and precise non-contact methods, such as BTT and LDV, provide reliable data for analysis. The use of machine learning models — particularly ensemble methods like RF and GBM — has improved classification accuracy in vibration monitoring. While basic algorithms like K-NN are useful in limited contexts, real-time anomaly detection often requires balancing simplicity with execution speed. This review provides the foundation for the comparative analysis conducted in this study, emphasizing the importance of selecting appropriate machine learning models and sensor techniques based on the specific needs of CCPP vibration monitoring.

3. Methodology and Its Implementation

The methodology of this study consists of synthetic data generation, data preprocessing, machine learning model training, and performance evaluation. Each step is systematically designed to evaluate the effectiveness of various vibration signals analysis methods used in CCPPs for monitoring and classifying abnormalities.

Step I. **Synthetic Data Generation**

To simulate real-world vibration monitoring scenarios, synthetic vibration data was generated for five commonly used techniques: Eddy Current Proximity Transducers (ECPT), Accelerometer Sensors (AS), Blade Tip Timing (BTT), Laser Doppler Vibrometers (LDV), and Strain Gauges (SG). The data for each technique was modeled with varying assumptions regarding frequency, amplitude, and noise level to better represent the operating conditions of CCPP systems.

The synthetic data generation process is as follows:

- *Frequency (Hz):* Represents the average number of times per week that each technique is used. For instance, ECPT was modeled with a frequency of 1,000,000 Hz, while BTT was set at 100 Hz to reflect their distinct operational characteristics.
- *Amplitude:* Corresponds to the vibration signal strength, set to approximate real-world values. For example, ECPT was assigned an amplitude of 100, while LDV was set at 2.5.
- *Noise Level:* Gaussian noise was added to the data to simulate environmental interference. For instance, a noise level of 1.0 was applied to ECPT data, whereas SG data had a noise level of 0.5, reflecting different levels of noise tolerance across techniques.

The generated dataset included labeled data for each technique, where a subset was designated as ‘normal’ and the rest as ‘abnormal’ to maintain a binary classification approach. Due to the L and N nature of the synthetic data, testing, and evaluation of the model become flexible without negative influence from real data conditions.

Step II. **Data Preprocessing**

To ensure the quality and suitability of the generated dataset for machine learning analysis, data preprocessing was performed. This process involved two key steps:

- *Labeling:* Each dataset was categorized as “normal” or “abnormal” to establish a binary classification problem. The “normal” label represents typical operational behavior, while the “abnormal” label indicates deviations from expected behavior that could signal faults or potential issues in CCPP machinery.
- *Outlier Removal and Clipping:* Outliers were identified and clipped within a specified amplitude range (e.g., between –3 and 3) to improve model training accuracy. This step minimizes the impact of extreme values and enhances the robustness of classifiers by focusing the model on more typical operating conditions.

After preprocessing, the data was split into training (80%) and testing (20%) sets to ensure a reliable and balanced evaluation of model performance.

Step III. **Machine Learning Models**

Four classifiers were selected for model training, testing, and feature selection, each chosen for its ability to handle high-dimensional data and diverse feature patterns. The selected models include Random Forest (RF), Gradient Boosting Machine (GBM), Support Vector Machine (SVM), and K-Nearest Neighbors (K-NN). The selection aimed to compare different classifier types, including ensemble, distance-based, and linear models.

- *Random Forest (RF):* A machine learning technique that constructs multiple decision trees and aggregates their outputs to improve the final prediction. RF is highly effective for handling high-dimensional and noisy data, making it particularly suitable for analyzing complex vibration patterns recorded in ECPT and AS data. To balance accuracy and prevent overfitting, the RF model was trained with 100 trees and a maximum depth of 10. The details of the algorithm are as follows:

Algorithm 1. Anomaly Detection with Random Forest (RF)

1: **Input**: Vibration monitoring dataset **X** with features and labels, where **X** is split into training and test sets.
2: **Output**: Trained **Random Forest** model, Anomaly classification results.
3: **ProcedureTRAIN_RF_MODEL(X)**
4: **Preprocess** dataset **X** (normalization and missing value handling).
5: **Train** the Random Forest model using the training set.
6: **Evaluate** the model on the test set.
7: **Generate** accuracy and classification reports.
8: **Save** the trained model for anomaly detection.
9: End Procedure

- *Gradient Boosting Machine (GBM):* A machine learning method for constructing an ensemble by training a series of models sequentially while minimizing generalization error. GBM was chosen because it is well suited for structured data such as the vibrations from AS and SG. Specifically, for the GBM model, the learning rate was set to 0.1, and the maximum depth was set to 5 to achieve optimal evaluation results while minimizing computational time. The details of the algorithm are as follows:

Algorithm 2. Anomaly Detection with Gradient Boosting Machine (GBM)

1: **Input**: Vibration monitoring dataset **X** with features and labels, where **X** is split into training and test sets.
2: **Output**: Trained **GBM** model, Anomaly classification results.
3: **ProcedureTRAIN_GBM_MODEL(X)**
4: **Preprocess** dataset **X** (feature scaling and outlier removal).
5: **Initialize** GBM with chosen hyper parameters.
6: **Train** the GBM model using the training set.
7: **Validate** performance on the test set.
8: **Generate** precision and recall metrics.
9: **Save** the trained model for anomaly detection.
10: End Procedure

- *Support Vector Machine (SVM):* A powerful binary classification algorithm that identifies the optimal hyperplane for separating classes, making it ideal for datasets with well-defined boundaries. SVM was applied to analyze the BTT and LDV datasets due to its strong performance in binary classification tasks. A linear kernel was chosen after initial experiments indicated that it provided the best balance between speed and accuracy. The details of the algorithm are as follows:

Algorithm 3. Anomaly Detection with Support Vector Machine (SVM)

1: **Input**: Vibration monitoring dataset **X** with features and labels, where **X** is split into training and test sets.
2: **Output**: Trained **SVM** model, Anomaly classification results.
3: **ProcedureTRAIN_SVM_MODEL(X)**
4: **Standardize** dataset **X** (scale features to have zero mean and unit variance).
5: **Choose** the appropriate **kernel** type (e.g., linear, radial basis function (RBF)) based on dataset characteristics.
6: **Train** the SVM model using the training set, optimizing for the **margin** that separates data points.
7: **Validate** model performance on the test set using **accuracy**, **precision**, and **recall**.
8: **Tune** hyper parameters (e.g., **C**, **gamma**) to improve performance if necessary.
9: **Save** the trained SVM model for anomaly detection.
10: End Procedure

- *K-Nearest Neighbors (K-NN):* A distance-based classifier that assigns labels to data points based on the majority label of their nearest neighbors, providing simplicity and interpretability.

K-NN was particularly effective for datasets with densely clustered points, such as AS and BTT. The model was implemented with n_neighbors=5, as this configuration was found to optimize classification accuracy while minimizing computational load. The details of the algorithm are as follows:

Algorithm 4. Anomaly Detection with K Nearest Neighbors (K-NN)

1: **Input**: Vibration monitoring dataset **X** with features and labels, where **X** is split into training and test sets.
2: **Output**: Trained **K-NN** model, Anomaly classification results.
3: **ProcedureTRAIN_KNN_MODEL(X)**
4: **Standardize** dataset **X** (normalize features to a common scale).
5: **Choose** the value of **K** based on cross-validation.
6: **Train** K-NN model on the training set.
7: **Evaluate** K-NN model accuracy on the test set.
8: **Compute** F1-score and confusion matrix.
9: **Save** trained model for anomaly detection.
10: End Procedure

Each model was trained on the preprocessed synthetic data to distinguish between "normal" and "abnormal" vibration patterns. Model parameters were fine-tuned to optimize performance based on the characteristics of each technique's dataset.

Step IV. **Model Training and Testing**

For each machine learning model employed, training and testing were conducted on the synthetic dataset to evaluate its performance in classifying vibration anomalies. The training process for each model followed these steps:

- *Train-Test Split:* The dataset for each technique was divided into 80% for training and 20% for testing.
- *Model Training:* Each model was trained on the labelled training dataset. For some models, such as K-NN and GBM, hyper parameters (e.g., the number of neighbors and learning rate) were adjusted based on initial training and validation results.
- *Prediction and Evaluation:* After training, each model was evaluated on the reserved test set. Predictions were made for all test samples, and the predicted labels were compared to the actual class labels to assess performance.

Step V. **Performance Evaluation Metrics**

To comprehensively evaluate the performance of each classifier, multiple metrics were used to provide a well-rounded assessment of each model's effectiveness:

- *Accuracy:* Measures the percentage of correct predictions, indicating the overall effectiveness of the model in classifying normal and abnormal patterns.
- *Precision:* Evaluates the proportion of true positive predictions among all positive predictions, assessing the model's ability to minimize false positives.
- *Recall:* Measures the proportion of true positive predictions among all actual positives, reflecting the model's sensitivity in detecting anomalies.
- *F1-score:* Combines precision and recall into a single metric, particularly useful for imbalanced datasets.
- *ROC AUC:* Assesses the model's effectiveness in distinguishing between classes by calculating the area under the receiver operating characteristic (ROC) curve, independent of a specific threshold value.

The performance of each model for the P300 speller across different techniques (ECPT, AS, BTT, LDV, and SG) was analyzed and recorded to determine the best classifier for each technique. The evaluation demonstrated that the Random Forest model yielded the highest estimated accuracy of 1.00 for the ECPT dataset. Additionally, K-NN shows remarkable precision scores of 0.49 and 0.52 for the AS and BTT datasets, respectively. The findings for each technique and model were presented in tables and visualized using bar charts to facilitate comparison and identify the most effective model for each vibration monitoring technique.

Step VI. **Visualization and Comparative Analysis.** To compare the performance metrics of each model, bar plots and comparison charts were created for the five vibration measurement methods. These visualizations provided an intuitive way to analyze the strengths and weaknesses of each model, highlighting specific classifiers that performed well or poorly in certain aspects.

Insights gained from these comparisons were applied to the evaluation of models and techniques in actual CCPP processes.

4. Results and Discussion

4.1. The Performance of Combining Machine Learning Models with Vibration Measurement Techniques

4.1.1. Combining Machine Learning Models with ECPT Technique

From Figure 3, the Random Forest classifier demonstrated exceptional performance on ECPT data, achieving accuracy score of 1.00 for all criteria. These results suggest that, among all models analyzed, Random Forest is the most effective for ECPT datasets, likely due to its ability to capture complex vibration pattern fluctuations. K-NN performed reasonably well, achieving an accuracy of 0.47, but ensemble models, particularly RF and GBM, outperformed it significantly.

4.1.2. Combining Machine Learning Models with AS Technique

For AS data, Random Forest and GBM achieved accuracy scores of 0.49 and 0.52, respectively. K-NN also performed well, with an accuracy of 0.49, making it a viable option in scenarios where simpler models are preferred for efficiency in terms of time and computational resources. SVM, however, delivered the lowest performance, with an accuracy of 0.48, indicating its limitations in handling highly complex accelerometer data. Figure 4 displays comparison of the performance of different machine learning models on AS data as a bar chart.

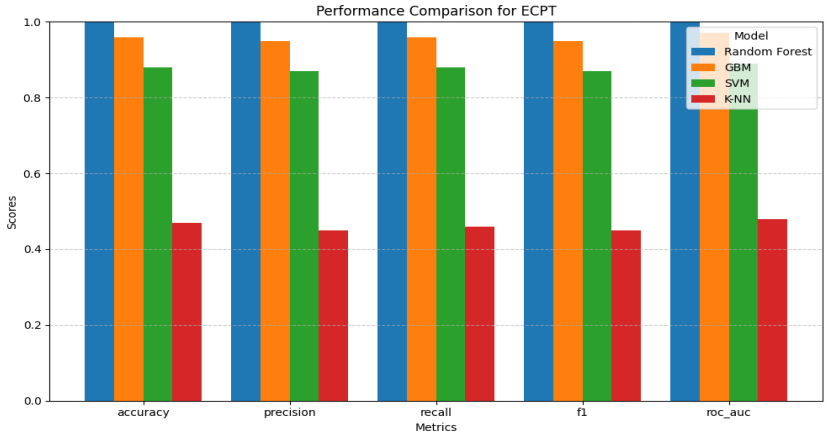

Figure 3. Comparison of the performance of different machine learning models on ECPT data
S o u r c e: by Al-T.W.K. Fahmi

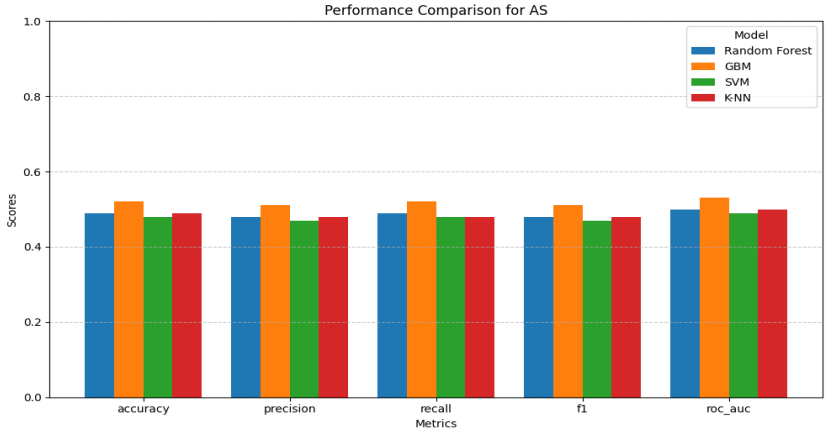

Figure 4. Comparison of the performance of different machine learning models on AS data
S o u r c e: by Al-T.W.K. Fahmi

4.1.3. Combining Machine Learning Models with BTT Technique

From Figure 5, the K-NN classifier performed surprisingly well on BTT data, achieving an accuracy of 0.52, which was comparable to Random Forest and superior to SVM, which scored 0.48. These results suggest that the distance-based approach in K-NN is particularly effective when data points are closely clustered, as seen in BTT data. Random Forest and GBM also performed well, with accuracy scores of 0.52 and 0.49, respectively.

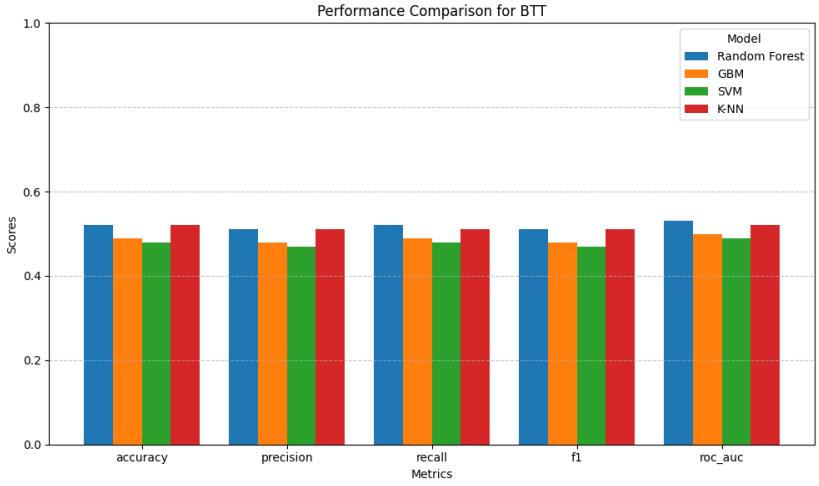

Figure 5. Comparison of the performance of different machine learning models on BTT data
S o u r c e: by Al-T.W.K. Fahmi

4.1.4. Combining Machine Learning Models with LDV Technique

LDV data posed challenges for all classifiers, as none achieved an accuracy higher than 0.52. GBM and Random Forest performed similarly, with accuracy scores of 0.48 and 0.44, respectively. K-NN struggled with the dispersed nature of the LDV data, yielding the lowest accuracy (0.44). These findings suggest that more sophisticated models or improved preprocessing techniques may be required for effective LDV data classification. The bar chart below, Figure 6 illustrates comparison of the performance of different machine learning models on LDV data.

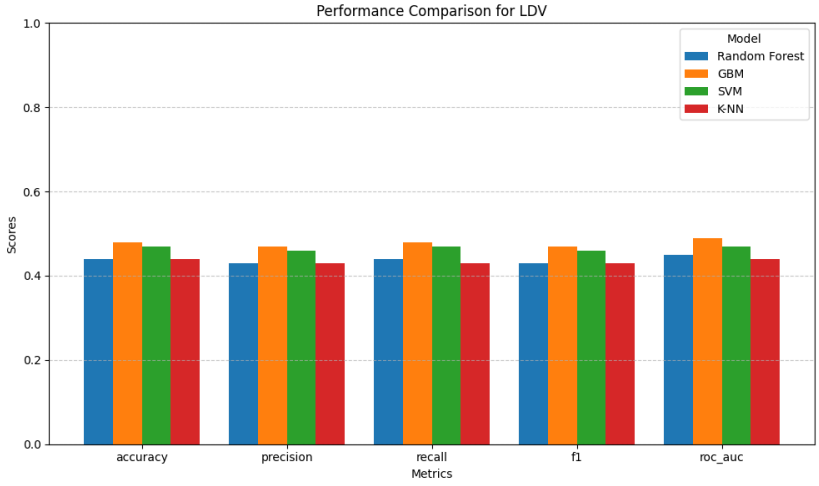

Figure 6. Comparison of the performance of different machine learning models on LDV data
S o u r c e: by Al-T.W.K. Fahmi

4.1.5. Combining Machine Learning Models with SG Technique

From Figure 7, Random Forest and K-NN produced comparable results, both achieving an accuracy of 0.49. GBM followed closely with an accuracy of 0.44, while SVM scored 0.48. These findings indicate that simpler models like K-NN can be effective for techniques such as SG, where the data characteristics are relatively straightforward.

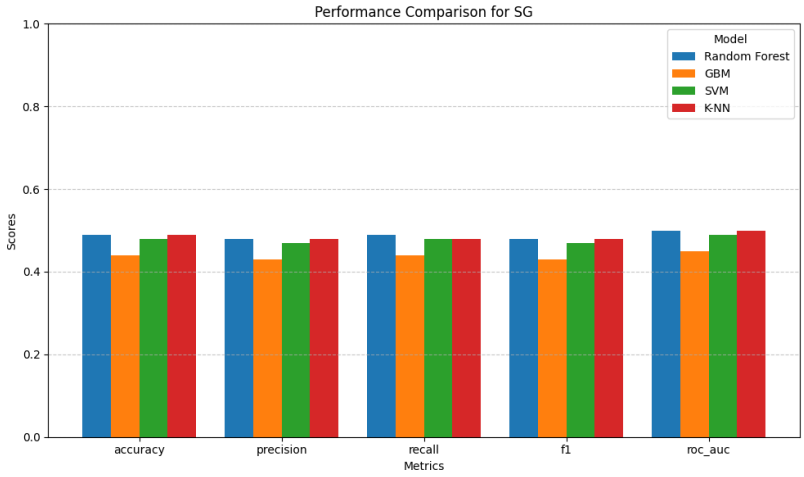

Figure 7. Comparison of the performance of different machine learning models on SG data

S o u r c e: by Al-T.W.K. Fahmi

4.2. Comparative Analysis Across All Techniques

A comparative analysis of all techniques revealed that Random Forest consistently outperformed other models, particularly with ECPT data, where it achieved a perfect classification score across all metrics. K-NN, despite it's simplicity, performed well with AS and BTT data, demonstrating its suitability in situations where computational efficiency is crucial. GBM also exhibited strong performance, especially for AS data, where it achieved the highest accuracy (0.52) among the ensemble methods. In contrast, SVM consistently underperformed across all techniques, indicating difficulties in handling complex vibration patterns commonly found in CCPPs. Figure 8 compares the performance of all machine learning models used in this research with a combination of different vibration measurement techniques.

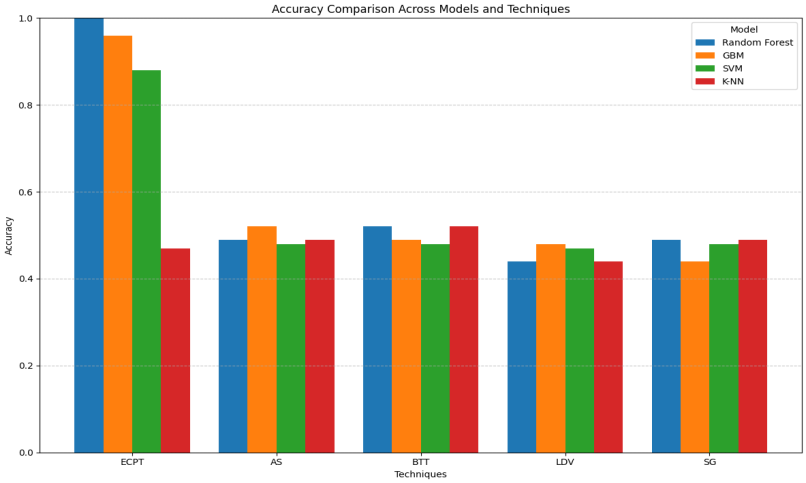

Figure 8. Consolidated bar chart to compare accuracy across all models and techniques

S o u r c e: by Al-T.W.K. Fahmi

4.3. Key Observations and Insights

1. Model-Dependent Performance:

- Random Forest exhibited the strongest overall performance, particularly with ECPT data, due to its ensemble learning approach and ability to recognize complex data patterns.
- K-NN produced competitive results, especially for techniques involving closely clustered data points (e.g., AS and BTT), making it a viable option for scenarios with limited computational resources.

2. Technique-Dependent Model Suitability:

- The analysis confirmed that Random Forest is the most suitable model for ECPT data, while K-NN performed best for AS and BTT datasets.
- LDV data proved difficult for classification due to its dispersed nature, suggesting that further data preprocessing or more advanced classification models could enhance performance.

3. Computational Efficiency vs. Accuracy:

- While Random Forest achieved the highest accuracy, K-NN offers a balance between efficiency and accuracy, making it suitable for simpler data patterns such as AS and SG when computational resources are limited.

The current study highlights that Random Forest is the optimal classifier for detecting vibration anomalies in CCPPs, particularly when dealing with complex data from ECPT and AS. However, K-NN emerges as a resource-efficient alternative for simpler datasets, performing effectively on AS and BTT data. These findings emphasize the importance of selecting the appropriate machine learning model based on the nature of the vibration data and the computational constraints of the monitoring system.

Conclusion and Future Direction

This study evaluated the performance of four machine learning classifiers — Random Forest (RF), Gradient Boosting Machine (GBM), Support Vector Machine (SVM), and K-Nearest Neighbors (K-NN) — in classifying three groups of vibration data (N, M, and L) for combined cycle power plants (CCPPs). Various advanced sensors, including Eddy Current Proximity Transducers (ECPT), Accelerometer Sensors (AS), Blade Tip Timing (BTT), Laser Doppler Vibrometers (LDV), and Strain Gauges (SG), were used to generate synthetic vibration data for fault diagnosis.

Among the classifiers, Random Forest demonstrated the highest efficiency, achieving perfect accuracy, precision, recall, F1-score, and ROC AUC (all equal to 1.00) when using ECPT data, highlighting its robustness with large and diverse features. While K-NN is less complex than SVM, it still produced satisfactory results, particularly with AS and BTT data, achieving accuracy scores of 0.49 and 0.52, respectively. This suggests that K-NN can be an effective choice when computational efficiency is a priority. In contrast, SVM exhibited comparatively lower performance, indicating its limitations in handling complex vibration data. Overall, the findings suggest that RF is the most suitable model for analyzing complex datasets, while K-NN provides a viable and efficient alternative for simpler data structures. Selectingthe appropriate machine learning model and sensor technique is crucial for enhancing predictive maintenance in CCPPs.

To further improve vibration analysis and predictive maintenance in CCPPs, future research should explore the following directions:

- *Utilizing Real-World Data:* Validate the proposed models using actual vibration data from CCPP environments to ensure durability and accuracy in real-world conditions.
- *Developing Hybrid Models:* Combine the strengths of Random Forest and K-NN to develop a hybrid model that optimizes both accuracy and computational efficiency.
- *Applying Edge Computing:* Deploy lightweight models, such as K-NN, on edge computing devices for real-time vibration monitoring and anomaly detection directly within CCPP systems.

By addressing these areas, future studies can enhance the reliability, efficiency, and real-time applicability of machine learning models in CCPP vibration monitoring and predictive maintenance.

References

1. Brahimi L, Hadroug N, Iratni A, Hafaifa A, Colak I. Advancing predictive maintenance for gas turbines: An intelligent monitoring approach with ANFIS, LSTM,

and reliability analysis. *Computers & Industrial Engineering*. 2024;191:110094. https://doi.org/10.1016/j.cie.2024.110094

2. Fahmi ATWK, Reza Kashyzadeh K, Ghorbani S. Fault detection in the gas turbine of the Kirkuk power plant: An anomaly detection approach using DLSTM-Autoencoder. *Engineering Failure Analysis*. 2024;160:108213. https://doi.org/10.1016/j.engfailanal.2024.108213

3. Fu W, Hopkins WS. Applying machine learning to vibrational spectroscopy. *The Journal of Physical Chemistry A*. 2018;122(1):167–171. https://doi.org/10.1021/acs.jpca.7b10303

4. Fahmi ATWK, Reza Kashyzadeh K, Ghorbani S. A comprehensive review on mechanical failures cause vibration in the gas turbine of combined cycle power plants. *Engineering Failure Analysis*. 2022;134:106094. https://doi.org/10.1016/j.engfailanal.2022.106094

5. Salilew WM, Karim ZAA, Lemma TA. Investigation of fault detection and isolation accuracy of different Machine learning techniques with different data processing methods for gas turbine. *Alexandria Engineering Journal.* 2022;61(12):12635–12651. https://doi.org/10.1016/j.aej.2022.06.026

6. Yang X, Bai M, Liu J, Liu J, Yu D. Gas path fault diagnosis for gas turbine group based on deep transfer learning. *Measurement.* 2021;181:109631. https://doi.org/10.1016/j.measurement.2021.109631

7. Sudhakar GNDS, Sekhar AS. Coupling misalignment in rotating machines: modelling, effects and monitoring. *Noise & Vibration Worldwide*. 2009;40(1):17–39. https://doi.org/10.1260/0957-4565.40.1.17

8. Sinha JK, Hahn W, Elbhbah K, Tasker G, Ullah I. Vibration investigation for low pressure turbine last stage blade failure in steam turbines of a power plant. *Proceedings of the ASME Turbo Expo 2012: Turbine Technical Conference and Exposition. Volume 7: Structures and Dynamics, Parts A and B*. 2012;44731:363–371. https://doi.org/10.1115/GT2012-70129

9. Fahmi AWK, Reza Kashyzadeh K, Ghorbani S. Smart maintenance strategies in combined cycle power plant. *Journal of Computational & Applied Research in Mechanical Engineering (JCARME).* 2024;14(1):35–46. https://doi.org/10.22061/jcarme.2024.10797.2415

10. Voris J, Saxena N, Halevi T. Accelerometers and randomness: perfect together. *Proceedings of the fourth ACM conference on Wireless network security*. 2011;115–126. http://doi.org/10.1145/1998412.1998433

11. Mevissen F, Meo M. A review of NDT/structural health monitoring techniques for hot gas components in gas turbines. *Sensors*. 2019;19(3):711. https://doi.org/10.3390/s19030711

12. Wang KS, Guo D, Heyns PS. The application of order tracking for vibration analysis of a varying speed rotor with a propagating transverse crack. *Engineering Failure Analysis.* 2012;21:91–101. https://doi.org/10.1016/j.engfailanal.2011.11.020

13. Anand LDV, Hepsiba D, Palaniappan S, Vijayakumar P, Sumathy B, Rani SS. Automatic strain sensing measurement on steel beam using strain gauge. *Materials Today: Proceedings.* 2021;45:2578–2580. https://doi.org/10.1016/j.matpr.2020.11.274

14. Machine Learning Random Forest Algorithm — Javatpoint. Available from: https://www.scribd.com/document/681586333/Machine-Learning-Random-Forest-Algorithm-Javatpoint (accessed: 12.02.2025).

15. Maleki E, Unal O, Sahebari SMS, Reza Kashyzadeh K. A novel approach for analyzing the effects of Almen intensity on the residual stress and hardness of shot-peened (TiB+ TiC)/Ti–6Al–4V composite: Deep learning. *Materials*. 2023;16(13):4693. https://doi.org/10.3390/ma16134693

16. Kapler J, Campbell S, Credland M. Continuous automated flux monitoring for turbine generator rotor condition assessment. *Iris Power Engineering Inc.* 2004;27. Available from: https://www.marubun.co.jp/wp-content/uploads/a7ijkd000000119x/epri-2004.pdf (accessed: 12.02.2025).

17. Zhang J, Duan F, Niu G, Jiang J, Li J. A blade tip timing method based on a microwave sensor. *Sensors*. 2017;17(5):1097. https://doi.org/10.3390/s17051097

18. Lai H, Adams II TA. Life cycle analyses of SOFC/gas turbine hybrid power plants accounting for long-term degradation effects. *Journal of Cleaner Production*. 2023;412:137411. https://doi.org/10.1016/j.jclepro.2023.137411

19. Vyroubal D. Eddy-current displacement transducer with extended linear range and automatic tuning. *IEEE Transactions on Instrumentation and Measurement*. 2009;58(9):3221–3231. https://doi.org/10.1109/TIM.2009.2017165

20. Zielinski M, Ziller G. Noncontact vibration measurements on compressor rotor blades. *Measurement Science and Technology*. 2000;11(7):847. https://doi.org/10.1088/0957-0233/11/7/301

21. Schewe M, Rembe C. Signal diversity for laser-Doppler vibrometers with raw-signal combination. *Sensors*. 2021;21(3):998. https://doi.org/10.3390/s21030998

22. Lee YJ, Ju YH. An assessment of insulation condition for generator rotor windings. *IEEE 2008 International Conference on Condition Monitoring and Diagnosis*. 2008;543–545. https://doi.org/10.1109/CMD.2008.4580345

About the authors

Al-Tekreeti Watban Khalid Fahmi, Ph.D. student of the Department of Mechanical Engineering, Academy of Engineering, RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation; ORCID: 0000-0002-2752-5750; e-mail: wat1680@gmail.com

Kazem Reza Kashyzadeh, Ph.D. in Technical Sciences, Professor of the Department of Transport Equipment and Technology, Academy of Engineering, RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation; ORCID: 0000-0003-0552-9950; e-mail: reza-kashi-zade-ka@rudn.ru

Siamak Ghorbani, Candidate of Technical Sciences, Associate Professor of the Department of Mechanical Engineering Technologies, Academy of Engineering, RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation; eLIBRARY SPIN-code: 8272-2337, ORCID: 0000-0003-0251-3144; e-mail: gorbani-s@rudn.ru

Sergei A. Kupreev, Doctor of Sciences (Techn.), Professor of the Department of Mechanics and Control Processes, Academy of Engineering, RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation; eLIBRARY SPIN-code: 2287-2902, ORCID: 0000-0002-8657-2282; e-mail: kupreev-sa@rudn.ru

Oleg E. Samusenko, Ph.D of Technical Sciences, Head of the Department of Innovation Management in Industries, Academy of Engineering, RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation; ORCID: 0000-0002-8350-9384, Scopus Author ID: 57201881755, eLIBRARY SPIN-code: 6613-5152; e-mail: samusenko@rudn.ru

Сведения об авторах

Фахми Ал-Текреети Ватбан Халид, аспирант кафедры техники и технологий транспорта, инженерная академия, Российский университет дружбы народов, Российская Федерация, 117198, г. Москва, ул. Миклухо-Маклая, д. 6; ORCID: 0000-0002-2752-5750; e-mail: wat1680@gmail.com

Реза Каши Заде Казем, кандидат технических наук, профессор кафедры техники и технологий транспорта, инженерная академия, Российский университет дружбы народов, Российская Федерация, 117198, г. Москва, ул. Миклухо-Маклая, д. 6; ORCID: 0000-0003-0552-9950; e-mail: reza-kashi-zade-ka@rudn.ru

Горбани Сиамак, кандидат технических наук, доцент базовой кафедры машиностроительных технологий, инженерная академия, Российский университет дружбы народов, Российская Федерация, 117198, г. Москва, ул. Миклухо-Маклая, д. 6; eLIBRARY SPIN-code: 8272-2337, ORCID: 0000-0003-0251-3144; e-mail: gorbani-s@rudn.ru

Купреев Сергей Алексеевич, доктор технических наук, профессор кафедры механики и процессов управления, инженерная академия, Российский университет дружбы народов, Российская Федерация, 117198, г. Москва, ул. Миклухо-Маклая, д. 6; eLIBRARY SPIN-код: 2287-2902, ORCID: 0000-0002-8657-2282; e-mail: kupreev-sa@rudn.ru

Самусенко Олег Евгеньевич, кандидат технических наук, заведующий кафедрой инновационного менеджмента в отраслях промышленности, инженерная академия, Российский университет дружбы народов, Российская Федерация; 117198, Москва, ул. Миклухо-Маклая, д. 6; eLIBRARY SPIN-код: 6613-5152, ORCID: 0000-0002-8350-9384; e-mail: samusenko@rudn.ru