Applied Computer Science 2026, vol. 22, no. 1, pp. 171–183
https://doi.org/10.35784/acs_7947

Submitted: 2025-06-18 | Revised: 2025-08-29 | Accepted: 2025-09-07

Keywords: AIoT, explainable AI, feature engineering, lubrication system, predictive maintenance

Ali KHUMAIDI [1*], *Risanto DARMAWAN* [2], *Lukman ADITYA* [2], *Wardhana Halking HAMKA* [3], *Hudzaifah Al JIHAD* [3]
[1] Universitas Nusa Mandiri, Indonesia, alikhumaidi.kmd@nusamandiri.ac.id
[2] Krisnadwipayana University, Indonesia, risantodarmawan@unkris.ac.id, lukmanaditya@unkris.ac.id
[3] PT. Multindo Technology Utama, Indonesia, wardhana.hamka@multindo-technology.com, hudzaifah@multindo-technology.com
[*] Corresponding author: alikhumaidi.kmd@nusamandiri.ac.id

Smart autolube: Optimized machine learning-based pressure prediction for AIoT lubrication systems

Abstract

Autolube systems have been widely adopted in the mining industry to improve equipment reliability, but most still operate at fixed time intervals without adapting to real conditions in the field, and monitoring systems use LED lights, making it difficult to diagnose failures due to the minimal system interface. To overcome these issues, this study developed Smart Autolube based on Artificial Intelligence of Things (AIoT), which integrates sensor-based monitoring with machine learning models for adaptive lubrication pressure prediction. With industry support from PT. Multindo Technology Utama, the system was tested under mining simulation conditions using pressure, temperature, and stress sensor data. After preprocessing, which includes winsorization, feature engineering (lag, rolling statistics, and trends), two ensemble algorithms, Random Forest Regressor (RFR) and Gradient Boosting Regressor (GBR), are used to build a prediction model. The base model showed low accuracy ($R^2 < 0.1$), but after feature engineering and extreme hyperparameter tuning, the performance improved significantly with an R^2 of 0.9816 for GBR and 0.9711 for RFR. Explanability analysis using SHAP (SHapley Additive exPlanations) shows that engineering features such as trends, lag_2, and rolling_mean_3 contribute the most to the predictions compared to native features such as temperature and voltage. This study proves that Smart Autolube can provide accurate and explainable lubrication pressure predictions. Further research is suggested to expand the scope of the data, add other mechanical parameters, and test the generalization of the model in different industrial environments.

1. INTRODUCTION

The mining industry is one of the sectors that relies heavily on the reliability of heavy machinery and equipment operating in extreme conditions, ranging from high loads and temperatures to exposure to contaminants such as dust (Dayo-Olupona et al., 2023). In such an operating environment, the lubrication system plays a critical role in maintaining the longevity and efficiency of mechanical components such as bearings, gears, and rotating shafts (Perremans et al., 2024). According to a report by the Society of Tribologists and Lubrication Engineers (STLE), proper lubrication can reduce unplanned downtime by up to 30% and extend equipment life by up to 50% (Society of Tribologists and Lubrication Engineers, 2023). Lubrication errors, such as overlubrication, underlubrication, and contamination, account for approximately 64% of all bearing failures, directly resulting in increased maintenance costs and production downtime (Hamran et al., 2020; Yin et al., 2024).

To address these challenges, automatic lubrication (autolube) systems have been widely adopted in the industry due to their ability to automatically distribute lubricants in measured quantities and at specific time intervals without the need for manual intervention (Peng et al., 2023; Sardhara & Tamboli, 2018). However, the majority of autolube systems in use today are still based on an open-loop approach, where lubrication is performed periodically based on a predetermined time, without considering actual operating conditions such as lubricant pressure, operating temperature, or stress stability. This limitation causes the system to act only as an automatic pump, without the ability to detect whether lubrication has actually occurred, whether pressure has been reached, or whether a system malfunction has occurred. In extreme field conditions, such as mining, this approach can result in lubrication that does not meet actual needs, such as lubricant pressure, operating temperature, or voltage stability, ultimately accelerating component wear, reducing energy efficiency, and increasing the risk of downtime due to undetected damage (Widarmadi et al., 2023; Wszelaczyński et al., 2021).

Mine operators in the field often find it difficult to determine the cause of Autolube system failure when a failure occurs. This is due to the limitations of the human-machine interface (HMI) commonly used in autolube equipment, which is limited to a single button and a red/green LED indicator. When the system fails, the LEDs simply indicate an error status without further explanation, often leading the operator to speculate that the autolube product is faulty, when the primary cause is usually spent lubricant, wiring problems or an improper input voltage. These limitations result in poor local diagnostic capabilities for operators, slow down the troubleshooting process, and have a direct impact on equipment downtime (Kalra et al., 2021; Meddeb et al., 2023). The use of simple interfaces without interactive diagnostic feedback has been shown to increase the risk of system misinterpretation, especially in stress-intensive work environments such as mining (Ma et al., 2019).

The development of Internet of Things (IoT) and artificial intelligence (AI) technologies has opened up significant opportunities to transform lubrication systems from traditional to smarter. IoT enables real-time data collection from various industrial sensors such as pressure, temperature, and voltage, while AI enables the analysis of such data to detect anomalies and formulate lubrication strategies based on actual conditions (Dayo-Olupona et al., 2023; Green & Taylor, 2024; Rahman et al., 2023). This approach has been shown to improve system reliability and reduce downtime in the context of predictive maintenance in manufacturing systems and heavy industrial machinery (Benhanifia et al., 2025; Henderson & Sanders, 2025). Recent research also highlights how the combination of IoT and AI in the form of Artificial Intelligence of Things (AIoT) can improve the adaptability of systems, making them suitable for real-time predictive maintenance in mining environments (Muhammed et al., 2024; Ucar et al., 2024). Research by Kumar et al. (2025) shows that the combination of sensors with algorithms such as ANN and ANFIS is able to predict lubricant life and system failures with excellent performance (correlation coefficient 0.99), opening up opportunities for the adoption of dynamic lubrication systems in mining and heavy industry. As a form of integration of the two approaches, the concept of Artificial Intelligence of Things (AIoT) has emerged as a framework that integrates data collection and analytical intelligence in a holistic manner (Awaisi et al., 2024).

Some researchers have attempted to address lubrication challenges through predictive modeling. Singh et al. (2022) demonstrated the potential of machine learning in optimizing lubrication and wear, while Nguyen Thanh & Cho (2024) explored a hybrid AIoT approach for anomaly conditional classification in industrial diesel generators. Similarly, Surucu et al. (2023) applied deep learning in predictive maintenance to minimize machine downtime and potential losses, while Liu et al. (2024) proposed machine learning-based fault diagnosis for several components of a hydraulic system. These studies highlight the growing relevance of AIoT in lubrication monitoring, but most of the existing research is still limited to general predictive maintenance without a strong focus on autolube systems.

A number of studies have explored the application of AI to the prediction of tribological parameters, including lubricant life and wear rate, but not many have explicitly targeted the prediction of lubrication pressure in autolube systems based on operational parameters such as temperature and voltage (Ardah et al., 2025; Rojas et al., 2025). Commercial solutions based on programmable logic controllers (PLCs) are available, but they are often expensive, complex, and do not meet the needs of mining customers, who are sensitive to cost, time efficiency, and limited technical resources (Sinitò et al., 2023). Therefore, a new approach that is more flexible, economical, and contextualized is needed.

To address this gap, this study developed an AIoT-based Smart Autolube that integrates pressure, temperature, and voltage sensors into a single real-time monitoring system and applies machine learning algorithms to predict lubrication pressure in an accurate and adaptive manner. The two models used and compared were the Random Forest Regressor and the Gradient Boosting Regressor. Random Forest is known for its ability to robustly handle multivariable data and is tolerant of overfitting (Wu et al., 2025), while Gradient Boosting excels at identifying highly accurate nonlinear patterns in energy and industrial forecasting systems (Krishnan et al., 2024; Val et al., 2024). By comparing the two models in the real-world context of an autolube system, this study aims to provide lubrication that is not only automatic, but also adaptive, efficient, and intelligent, thus driving the transformation from a time-based lubrication system to a predictive condition-based lubrication system that is practical and applicable in a modern mining and manufacturing environment that demands high reliability and cost efficiency.

2. SYSTEM DESIGN AND ARCHITECTURE

In order to overcome the limitations of conventional autolube, a smart autolube system was built in this study. The system architecture consists of five main components (Figure 1), namely (1) sensor layer, (2) edge processing, (3) communication, (4) IoT simulation and validation, and (5) cloud application. Each component is designed to support accurate lubrication decisions based on the actual pressure, temperature, and voltage conditions of the autolube system. The Smart Autolube system designed in this study has a number of strategic advantages. The modular architecture allows for flexibility in integration and deployment scenarios. The system can also operate independently, both online (via the cloud) and offline (locally at the edge). Its main advantage lies in its intelligent and adaptive function, i.e. the system is not only responsible for pumping lubricant, but also for understanding conditions, predicting condition-adaptive lubrication needs and automatically making corrections. This architecture ensures that the proposed solution is not only technically robust, but also practical, scalable and economically viable for industrial implementation.

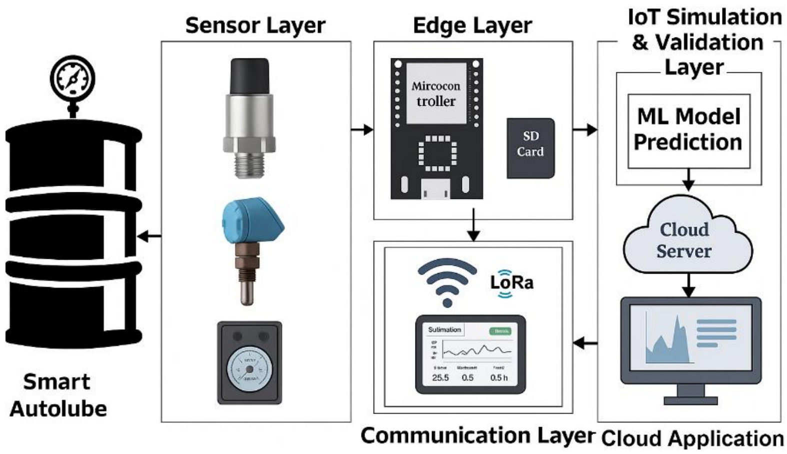


Fig. 1. Architecture of smart autolube

2.1. Sensor layer

The sensor layer consists of the main components responsible for the continuous acquisition of the functional parameters of the system. These sensors are installed at strategic points along the lubrication line and include (1) Pressure sensors to measure the actual pressure in the lube line. (2) Temperature sensors to monitor the thermal condition of the system and (3) Voltage sensors to ensure the stability of the input power to the Autolube pump. The modular design of the sensor layer also allows for future integration of additional parameters, such as lubricant flow rate or pump vibration, to expand the data set and improve prediction accuracy.

2.2. Edge processing layer

The data generated by the sensor is collected and processed locally by a microcontroller. These microcontrollers not only serve as the initial data collector and processor, but also store data locally on the SD card module as historical logs. To facilitate operator interaction in the field, the system is equipped with a touchscreen-based human-machine interface (HMI) that allows live monitoring, setup and repair. Edge processing also ensures computational efficiency by allowing essential pre-processing and anomaly detection to be performed locally in real time before data is sent to the cloud.

2.3. Communication layer

The communication layer is responsible for sending data from the edge device to the cloud server. The system supports Wi-Fi connectivity for locations with adequate network infrastructure and LoRa for field conditions with network limitations.

2.4. IoT simulation and validation layer

To increase reliability, an IoT-based simulation and validation layer is introduced before data is sent to the cloud. This layer allows scenario tests such as pressure loss, voltage fluctuations, and temperature variations to be validated against expected model behavior. By integrating simulation with live data, the system can distinguish between random noise and true anomalies, improving the robustness of predictions.

2.5. Cloud application layer

On the cloud-based application side, the received data is automatically compiled and analyzed using machine learning models. Predictive models are used to evaluate actual lubrication conditions and provide adaptive recommendations based on optimal pressure estimates. The system detects potential anomalies, such as pressure drops due to leaks or pump failures, and notifications are automatically sent to the cloud dashboard and displayed on the HMI.

3. MATERIALS AND METHODS

This research was conducted in collaboration with PT. Multindo Technology Utama (MTU) as a strategic industry partner. The company is a manufacturer of the commercial autolube product "Multilube" as an object of research and facilitates the testing and validation of the system in the field. The research process began with the assembly and integration of sensors on the Multilube unit, development of the HMI and cloud configuration. After the integration process was completed, sensor calibration and testing were performed to ensure the accuracy of data acquisition. To increase experimental reliability, a structured protocol was applied, including calibration before and after each experiment, verification of logging synchronization, and data integrity checks prior to modeling. The main stages in the development of Smart Autolube are (1) data acquisition, (2) preprocessing, (3) modeling, (4) model evaluation, and (5) explainable AI, as shown in Figure 2.

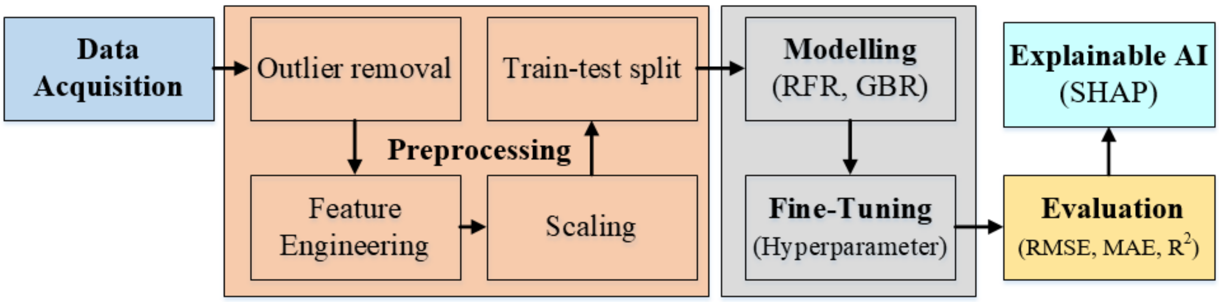


Fig. 2. Research stages of smart autolube

3.1. Data acquisition

The data was collected at the facilities of PT. MTU Sukabumi Branch, Indonesia, which replicates the operating conditions of the mining industry. The data collection process involves running a prototype for 5 days with a 10-hour workday and data collection every 5 minutes. Data is collected from three types of sensors and logged to a microSD card, which periodically sends data to the cloud. Each data collection is timestamped to ensure time synchronization (Sanap, 2025). To minimize bias, data sets were obtained under varying load conditions and ambient temperatures to ensure representation of both normal and stressed lubrication conditions.

3.2. Data preprocessing

Before being used in modeling, the data collected by the sensor is processed in several steps, namely outlier cleaning, which is performed using the Winsorization method to suppress extreme values that can distort the data distribution by limiting the maximum and minimum values based on a certain percentile, thus maintaining the sample size but reducing the effect of outlier values (Mekelburg & Strauss, 2024). In addition, feature engineering aims to enrich the representation of data by adding derivative features that are temporal. These features include lag features (pressure at previous time), rolling mean, rolling standard deviation, and trend features calculated from the difference between the current and previous pressures (Daraghmeh et al., 2025). These features are created to capture historical patterns and fluctuations in lubrication pressure that cannot be represented by native variables alone (such as temperature and stress).

After the feature is created, a separation process is performed between the feature (X) and the target (y), using the lubrication pressure as the prediction target. Furthermore, the data on the feature is normalized using the min-max scaling method, with the aim of placing all variables in the range [0, 1], so that the machine learning model can operate more stably and converge faster, especially for ensemble-based algorithms. Finally, the dataset is split into training and test data in an 80:20 ratio using a time-aware split approach. This method preserves the temporal order of the data so that the model does not violate temporal assumptions and is more representative in mimicking predicted conditions in the field (Al-Hares et al., 2021).

3.3. Modelling

Two regression algorithms are used to build a pressure prediction system, the Random Forest Regressor and the Gradient Boosting Regressor. Both are ensemble algorithms that can efficiently handle nonlinear and multivariate data and have tolerance for noise and overfitting when appropriately configured.

3.3.1. Random Forest Regressor (RFR)

RFR is a decision tree-based ensemble method that combines the results of multiple randomly constructed trees using the bootstrap aggregation (bagging) technique. This algorithm generates regression predictions by calculating the average predictions of all the trees constructed. RFR excels at reducing overfitting and efficiently handling multivariate data, making it particularly effective for small- to medium-sized datasets with complex variable structures (Probst et al., 2019). For the RFR prediction function, if there are N regression trees f1(x), f2(x), ... fN(x), then the final prediction (y) for input x can be seen in Eq. (1).

$$y = \frac{1}{N}\sum_{i=1}^{N} f_i(x) \tag{1}$$

The main advantage of RFR is its ability to reduce overfitting and produce stable estimates even for data with intercorrelated input variables (Q. Zhang et al., 2024). However, the performance of the model is strongly influenced by the configuration of the hyperparameters. Therefore, to ensure optimal generalization, the extreme hyperparameter tuning process was performed using the randomized search method in the following parameter spaces: n_estimators, max_depth, min_samples_split, min_samples_leaf, max_features, and bootstrap (Liang et al., 2024; Wang et al., 2023).

3.3.2. Gradient Boosting Regressor (GBR)

GBR is a sequential decision tree-based ensemble method in which each new model is trained to correct the prediction errors of previous models. This method optimizes the loss function by a gradient descent approach in function space (W. Zhang et al., 2024). The final model is expressed as the sum of several weak learners. The GBR prediction function, if there are N successive regression trees f1(x), f2(x), ... fN(x), then the final prediction (y) for input x is obtained by a stepwise sum of the contributions of each tree with a learning rate scale γ as shown in equation (2).

$$y = \sum_{n=1}^{N} \gamma_n f_n(x) \tag{2}$$

Like RFR, GBR is very sensitive to parameter configuration. Therefore, extensive fine-tuning was performed on the following hyperparameters: learning_rate, n_estimators, max_depth, min_samples_split, min_samples_leaf, subsample, and max_features. The tuning process is performed using the randomized search method, which allows efficient exploration of a large parameter space while preventing overfitting through cross-validation (Khan et al., 2025).

The extreme fine-tuning approach finds the optimal combination of parameters based on the model performance on the validation set and the results of the regression metric evaluation (Yakoubi et al., 2023). The optimization results show a significant improvement in prediction accuracy compared to the default configuration. This deeper parameter tuning is critical given the dynamic and non-linear nature of the lubrication data.

3.4. Model evaluation

To improve generalization, cross-validation (k=5) was performed in the training phase. Model evaluation was performed using three regression metrics, namely root mean square error (RMSE), mean absolute error (MAE), and R-squared (R^2). The RMSE measures the mean of the square root of the difference between the value predicted by the model and the actual value, and the MAE measures the mean of the absolute value of the difference between the prediction and the actual value. The smaller the RMSE and MAE values, the better the performance of the model. R-squared, or coefficient of determination, measures the proportion of variance in the target data that can be explained by the predictors in the model (Khumaidi et al., 2022). Using a combination of these three metrics ensures a comprehensive evaluation that balances error size with explanatory power. The value ranges from $-\infty$ to 1. The higher the value, the better the performance of the model.

3.5. Explainable artificial intelligence

This study uses the Explainable Artificial Intelligence (XAI) approach with the SHapley Additive exPlanations (SHAP) method to explain the contribution of each feature to the output of the pressure prediction model (Darvishvand et al., 2025).This XAI layer was particularly important for industrial adoption because it allowed engineers to validate that the learned patterns of the model were consistent with domain expertise, thus bridging AI predictions with human decision making. SHAP works by calculating the marginal contribution of each feature based on the principles of Shapley value game theory, allowing visualization of how much influence each feature (e.g., temperature, voltage, delay, rolling statistics, or trends) has on the predicted outcome. Through SHAP graphs, the model becomes more transparent and interpretable, making it easier for users to understand the reasoning behind the resulting predictions (Chen et al., 2024).

4. RESULTS AND DISCUSSION

4.1. Statistical results

The first step in data analysis is to check the distribution and consistency of the data using descriptive statistics. Table 1 shows the statistical parameters of the three observed sensors, namely pressure, voltage and temperature, based on a total of 720 data sets obtained from the smart autolube system during the simulation period.

Tab. 1. Descriptive statistics dataset

Number	Pressure	Voltage	Temperature	Timeupdate
Count	720	720	720	720
Mean	114.5183	25.3552	95.9298	2024-12-21 00:57:30
Min	61.3400	24.0200	39.1700	2024-12-18 08:00:00
25%	95.6625	25.1100	91.4800	2024-12-19 12:58:45
50%	114.4800	25.3300	98.3000	2024-12-21 00:57:30
75%	133.0875	25.5700	105.3450	2024-12-22 12:56:15
Max	149.9800	26.9900	120.5600	2024-12-23 17:55:00
Std	21.6092	0.4543	15.1819	-

The average pressure value is 114.52 psi, with a maximum of 149.98 psi and a minimum of 61.34 psi, indicating a fairly large variation (standard deviation of 21.61 psi). This indicates that the Autolube system is indeed subject to significant fluctuations, which may reflect real-world operating dynamics such as intermittent lubrication cycles or varying mechanical loads. The temperature ranges from 39.17°C to 120.56°C with an average of 95.93°C, which represents the autolube condition when the temperature is still low at the start of operation and quite high when the temperature is normal. Meanwhile, the voltage is relatively stable with a range of 24.02V to 26.99V, it has a low standard deviation of 0.45V, which indicates the stability of the power supply in the system.

The histogram visualization (Figure 3) shows that the pressure and temperature distributions tend to be close to normal, although there is a slight asymmetry in the temperature. These visualizations also help to identify potential outliers, which were treated during the preprocessing stage using the winsorization method, thus ensuring the reliability of the data for modeling. In addition, a Pearson correlation analysis was performed to evaluate the linear relationship between the input variables and the main target, namely pressure. The results of the analysis showed a very weak correlation value: temperature = 0.01 and voltage = -0.03 to pressure. These results indicate that both the grazing and the voltage do not have a significant linear relationship with the pressure value. This correlation analysis is also illustrated by a correlation matrix (Figure 4).

However, the low linear correlation does not mean that the two variables are irrelevant. This phenomenon leads to the possibility of nonlinear relationships that cannot be explained by Pearson correlation, making the use of machine learning algorithms such as Random Forest Regressor (RFR) and Gradient Boosting Regressor (GBR) relevant. Both models are capable of capturing complex patterns hidden in nonlinear data and are described in more detail in the next subsection. These statistical analyses ensure that the input data is reliable and representative of real-world operating conditions. By confirming data quality and detecting hidden nonlinear patterns, subsequent predictive models can provide more accurate forecasts, supporting reliable lubrication scheduling and minimizing unexpected downtime in heavy equipment operations.

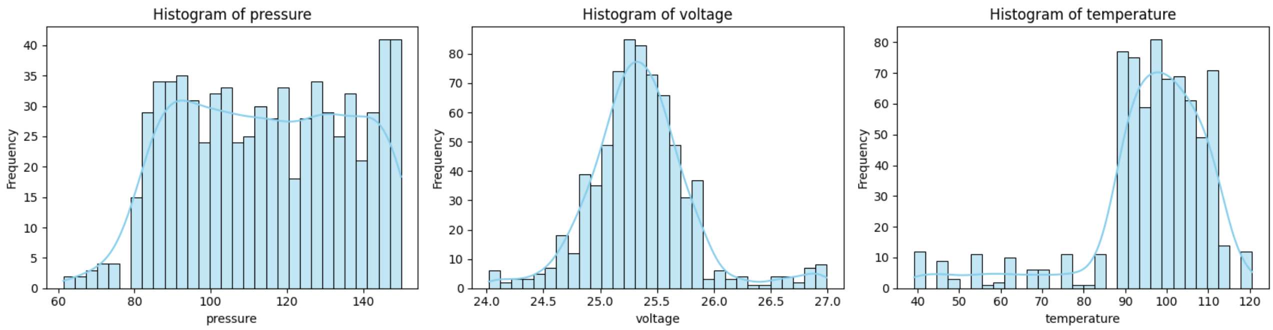


Fig. 3. Histogram of each parameter test

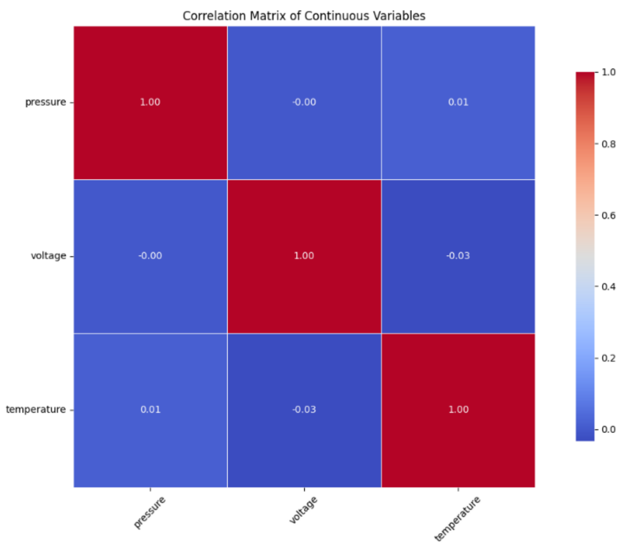


Fig. 4. Correlation matrix in autolube dataset

4.2. Baseline model

To build a machine learning-based predictive model, the dataset is split into two subsets: 80% (training set) and 20% (test set), using a time-aware split approach to preserve the temporal order of the data. This separation aims to ensure that the model has good generalization capabilities and does not overfit the training data. The two base models used in this study are RFR and GBR. Training of both models was performed using the default hyperparameters of the scikit-learn library, without initial optimization, to provide a baseline performance picture of each algorithm.

Tab. 2. Results of the evaluation of the basic model

Model	RMSE	MAE	R^2
GBR	18.84	15.98	0.088
RFR	18.82	15.41	0.090

The results of the model evaluation are shown in Table 2, where the RFR model produces an RMSE value of 18.82 and an MAE of 15.41, slightly better than the GBR. The value of the coefficient of determination (R^2) for both models was also relatively low, namely 0.090 for RFR and 0.088 for GBR, indicating that the baseline model was only able to explain less than 10% of the variability in the target data. The high error values and low predictive performance of these two models indicate that the initial input parameters (temperature, stress, error, and override) are not sufficiently representative to explain the pressure dynamics both linearly and nonlinearly. This weak performance is not unexpected since only raw features were provided, without temporal dynamics. In an industrial context, such weak baselines highlight the importance of domain-driven feature enrichment. This condition promotes the need for feature engineering to extract additional attributes capable of capturing temporal patterns and more complex system dynamics to improve prediction performance.

4.3. Feature engineering and extreme fine-tuning

The result is a linear correlation between features of a sensor with a very low target (Pearson's correlation). This indicates that the relationship pattern between input features and targets is nonlinear, so feature enrichment efforts are needed to capture temporal dynamics and pressure patterns hidden in time. To overcome these limitations, the feature engineering phase was carried out with a time series approach. Several feature transformations are implemented to form new features that can improve the model's ability to represent pressure patterns, namely.

- Lag Features: Represents the value of the pressure at the previous time. This feature is created by shift() of pressures 1, 2, 3, 6, and 12 periods back (lag_1, lag_2, ..., lag_12) with the goal that the model can learn the relationship between the current pressure and the previous pressure.
- Rolling Statistics: Mean and standard deviation of the pressure in rolling windows (rolling_mean and rolling_std) with windows 3, 5, and 12 with the goal of capturing information about pressure fluctuations and system stability over short to medium time periods.
- Trend Feature: The difference between the current pressure and the previous pressure (trend), which represents the immediate change in the direction of the pressure. This feature helps the model understand the tendency of the pressure to increase or decrease.

All of these features are added to the dataset, then a dropna() process is performed to eliminate the starting line that does not have enough history, as well as resetting the index to maintain the data structure. After feature engineering, the model was retrained using RFR and GBR. The results of the model performance evaluation are shown in Table 3.

Tab. 3. Model evaluation results after feature engineering

Model	RMSE	MAE	R^2
GBR	10.9648	9.1613	0.6928
RFR	3.5368	2.8363	0.9680

This result shows a very significant improvement in performance compared to the base model. The R^2 increased from <0.1 to 0.6928 for GBR and 0.9680 for RFR, indicating that the additional features were able to significantly enrich the input information and allow the model to better understand the pressure dynamics.

The decrease in RMSE and MAE values also reflects an increase in prediction accuracy on an absolute scale. Thus, feature engineering proves to be a crucial step in this AIoT-based prediction system, especially in the context of time signals such as pressure on the autolube system, which is highly influenced by previous conditions and short-term fluctuations.

The extreme fine-tuning stage aims to optimize model performance by searching for the best combination of hyperparameters using the randomized search approach. Each configuration is tested on training data. The results of the model evaluation are shown in Table 4.

Tab. 4. Results of model evaluation after extreme fine-tuning

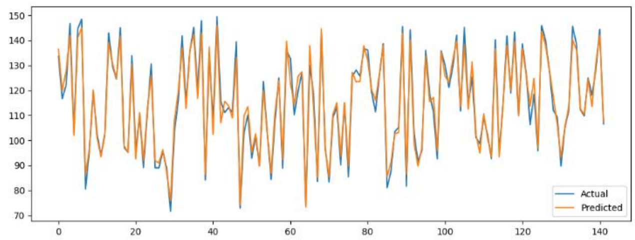
Model	RMSE	MAE	R^2
GBR	2.6823	2.1572	0.9816
RFR	3.3647	2.7379	0.9711

These results show that the fine-tuned GBR produces the best performance, with an R^2 value of 0.9816, indicating very high predictive ability. The RMSE and MAE values also decreased significantly compared to the previous model, reflecting a substantial improvement in prediction accuracy. This increase in accuracy is achieved through the selection of more optimal combinations of learning_rate and max_depth hyperparameters, as well as the use of subsamples that maintain model diversity. In the RFR model, max_features and min_samples_leaf parameters were shown to affect the stability and generalization of the model. Although the number of parameters and the complexity of the model increases after fine-tuning, it provides commensurate results in the form of better generalizability, higher accuracy, and greater adaptability to real data variations.

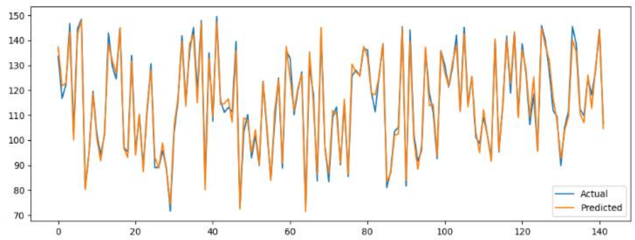
The results obtained in this study are significantly superior when compared to previous works. Kurnianto et al., (2024) show that linear regression provides performance with an R^2 of 0.772 for engine failure prediction. Similarly, Iftikhar et al. (2022) reported that support vector regression and SGDRegressor underperformed engine maintenance predictions with R^2 values of -2.428 and 0.393. In contrast, our proposed approach with RFR and GBR after feature engineering and fine-tuning reached an R^2 value of up to 0.9816, showing a substantial improvement in predictive reliability. It highlights that tree-based ensemble methods are better able to model complex nonlinear relationships and temporal patterns in autolube pressure prediction, thus offering a stronger foundation for AIoT-based predictive maintenance applications.

Figure 5 shows a comparison between the actual value and the pressure prediction results on autolube generated by the two algorithms. Both graphs show that visually, both models are able to follow a general pattern of pressure fluctuations, although there are still deviations at some observation points. The overall prediction distribution pattern that is quite close to the actual value indicates that both models have early capabilities in capturing pressure dynamics. The closer alignment of GBR predictions to actual data suggests its robustness against local fluctuations, which is essential in real-world deployment where operating conditions can shift.

In addition to accuracy improvements, computational efficiency was also considered for real-time deployment. While the fine-tuned GBR model achieved the highest accuracy, its higher complexity may increase computational load compared to RFR. This trade-off is important in industrial applications where prediction must be performed in real time with minimal latency. Both models, however, remain suitable for implementation on edge devices, which can process sensor data locally without relying on cloud infrastructure. Such deployment ensures faster response, reduces network dependency, and supports the scalability of the smart autolube system in real mining environments.



a) Actual vs Predicted Pressure of Random Forest Regression



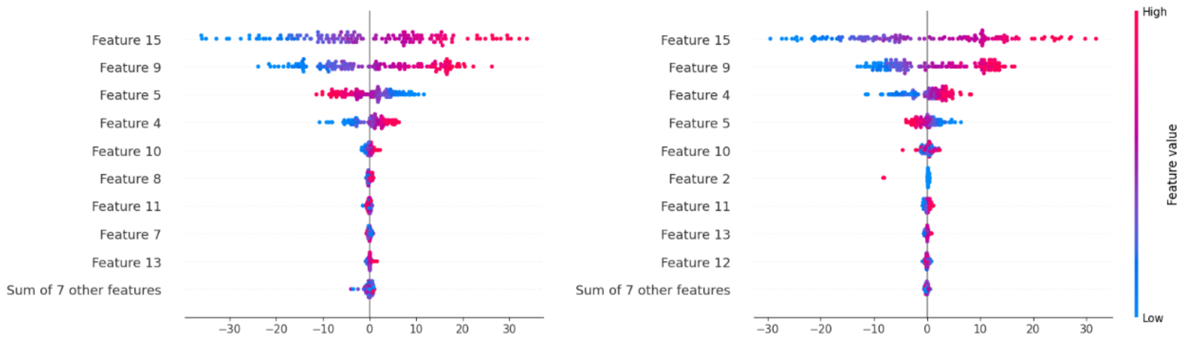
b) Actual vs Predicted Pressure of Gradient Boosting Regression

Fig. 5. Comparison chart of actual and predicted pressure

4.4. Explainable artificial intelligence

Explainable AI (XAI) analysis to understand the contribution of each feature to the lubrication pressure prediction generated by the two models. Figure 6 shows a SHAP (SHapley Additive exPlanations) plot showing the impact of each feature on the model output based on the SHAP value. This graph is very important because it provides a transparent interpretation of the model being used, so that it can be understood which features have the most impact on the prediction results. In Figure 7a (Random Forest Regression model), we can see that feature 15 (trend), feature 9 (rolling_mean_3), and feature 5 (lag_2) have the most significant contribution to the model's output. Meanwhile, Figure 7b (Gradient Boosting Regression model) shows a relatively similar pattern, with feature 15 (trend), feature 9 (rolling_mean_3), and feature 4 (lag_1) as the main contributors to the prediction process. Both models show consistency that features resulting from the feature engineering process, such as trend, lag, or rolling mean, are more dominant than the original features (such as voltage or temperature). This demonstrates the effectiveness of feature engineering strategies in improving model performance and better explaining the model's decision-making process and confirms that the system's predictive power lies in its ability to use historical patterns, not just instantaneous readings.

Both RFR and GBR showed consistent SHAP patterns, with "trend" and "rolling_mean_3" as the top contributors. For industrial practitioners, this means that engineers can interpret pressure predictions not as black-box outputs, but as the result of meaningful, explainable temporal signals. Overall, the SHAP analysis and prediction curves confirm that the AIoT-based pressure prediction model is not only accurate ($R^2 > 0.98$), but also interpretable. This approach not only provides high prediction accuracy but also improves the reliability of the Autolube system in terms of diagnostics and data-driven decision making. This dual benefit (accuracy + interpretability) strengthens the system's industrial readiness.



a) SHAP of Random Forest Regression   b) SHAP of Gradient Boosting Regression

Fig. 6. Diagram SHAP system autolube

5. CONCLUSIONS

This research successfully develops and implements an AIoT-based smart autolube system that integrates pressure, temperature, and voltage sensors with advanced machine learning models, namely RFR and GBR. Through the approaches of feature engineering, extreme fine tuning, and explainable artificial intelligence, the system is not only able to predict lubrication pressure in real time with high accuracy (R^2 to 0.9816) but also provides a transparent interpretation of the contribution of features to the prediction results using the SHAP method. The results of the experiment show that the trend features, the pressure lag (lag_1 and lag_2) and the rolling statistics (mean and standard deviation) contribute significantly to the improvement of the model performance. The SHAP value also shows that these features consistently positively affect the prediction results, supporting the reliability of condition-adaptive lubrication system decisions.

The practical implications of this study are particularly relevant in the context of heavy industry, especially mining, where downtime due to lubrication failures can result in large losses. By integrating this system, operators can optimize maintenance schedules and improve machine life. However, the study has several limitations, including the limited number of datasets (720 samples), the flexibility of AIoT integration and hyperparameter tuning can increase the computational load, and focusing only on pressure parameters as the prediction target may ignore other variables such as lubricant flow, mechanical condition of the pump, or vibration. These limitations indicate the need for larger, more diverse data sets and broader sensor integration to ensure robustness and scalability. Further research is suggested to develop datasets with a wider range of time and condition variations, integrate additional variables such as flow rate, pump condition, filter status, or vibration test model generalizations in other mine environments with different autolube configurations, and improve the system's diagnostic capabilities to generate corrective action recommendations based on explainability.

Beyond the technical contributions, the proposed system also has significant economic implications. By reducing unexpected downtime and minimizing lubrication-related failures, the Smart Autolube system can significantly reduce direct maintenance costs and improve overall equipment availability. In heavy mining operations, where machine downtime translates into significant financial losses, the use of this AIoT-based predictive maintenance framework offers not only technical reliability, but also tangible cost savings and improved return on investment. These economic benefits strengthen the value proposition of the system and support its adoption in industrial practice. In the long term, this framework can be expanded into a comprehensive AIoT-based predictive maintenance platform that not only monitors but autonomously adjusts lubrication strategies, creating a more resilient and intelligent mining ecosystem.

Funding

This research was funded by the Directorate General of Higher Education, Research, and Technology of the Republic of Indonesia and PT. Multindo Technology Utama through the Matching Fund Program scheme with contract number 8/E1/PPK/KS.03.00/2024.

Conflicts of interest

The authors declare no conflict.

REFERENCES

Al-Hares, M. K., Assimakopoulos, P., & Gomes, N. J. (2021). Ethernet fronthaul transport of an upper-physical layer functional split using time-aware shaping. *Optical Switching and Networking*, *41*, 100605. https://doi.org/10.1016/j.osn.2021.100605

Ardah, S., Profito, F. J., & Dini, D. (2025). A comprehensive review and trends in lubrication modelling. *Advances in Colloid and Interface Science*, *342*, 103492. https://doi.org/10.1016/j.cis.2025.103492

Awaisi, K. S., Ye, Q., & Sampalli, S. (2024). A survey of industrial AIoT: Opportunities, challenges, and directions. *IEEE Access*, *12*, 96946–96996. https://doi.org/10.1109/ACCESS.2024.3426279

Benhanifia, A., Cheikh, Z. Ben, Oliveira, P. M., Valente, A., & Lima, J. (2025). Systematic review of predictive maintenance practices in the manufacturing sector. *Intelligent Systems with Applications*, *26*, 200501. https://doi.org/10.1016/j.iswa.2025.200501

Chen, Y., Chen, B., & Shayilan, A. (2024). Combining categorical boosting and Shapley additive explanations for building an interpretable ensemble classifier for identifying mineralization-related geochemical anomalies. *Ore Geology Reviews*, *173*, 106263. https://doi.org/10.1016/j.oregeorev.2024.106263

Daraghmeh, M., Jararweh, Y., & Agarwal, A. (2025). Leveraging machine learning and feature engineering for optimal data-driven scaling decision in serverless computing. *Simulation Modelling Practice and Theory*, *140*, 103090. https://doi.org/10.1016/j.simpat.2025.103090

Darvishvand, L., Kamkari, B., Huang, M. J., & Hewitt, N. J. (2025). A systematic review of explainable artificial intelligence in urban building energy modeling: methods, applications, and future directions. *Sustainable Cities and Society*, *128*, 106492. https://doi.org/10.1016/j.scs.2025.106492

Dayo-Olupona, O., Genc, B., Celik, T., & Bada, S. (2023). Adoptable approaches to predictive maintenance in mining industry: An overview. *Resources Policy*, *86*, 104291. https://doi.org/10.1016/j.resourpol.2023.104291

Green, E., & Taylor, L. (2024). Development of a predictive maintenance framework for hydraulic systems using iot and machine learning. *International Journal of Mechanical, Industrial and Control Systems Engineering*, *1*(2), 48–53. https://doi.org/10.61132/ijmicse.v1i2.83

Hamran, N. N. N., Ghani, J. A., Ramli, R., & Haron, C. H. C. (2020). A review on recent development of minimum quantity lubrication for sustainable machining. *Journal of Cleaner Production*, *268*, 122165. https://doi.org/10.1016/j.jclepro.2020.122165

Henderson, J., & Sanders, M. (2025). AI driven predictive maintenance: reducing downtime and enhancing productivity in manufacturing environments. https://doi.org/10.20944/preprints202504.0602.v1

Iftikhar, N., Lin, Y.-C., & Nordbjerg, F. (2022). Machine learning based predictive maintenance in manufacturing industry. *3rd International Conference on Innovative Intelligent Industrial Production and Logistics* (pp. 85–93). SciTePress. https://doi.org/10.5220/0011537300003329

Widarmadi, I., Anam, S., Wisnu, B., & Sumardi, H. (2023). Analysis of the auto lube system electric grease pump against the unscheduled breakdown of the dump truck unit brand k type HD 1500-7. *Jurnal Terapan Teknik Mesin*, *4*(2), 197–206. https://doi.org/10.37373/jttm.v4i2.620

Kalra, J., Lightner, N. J., & Taiar, R. (2021). Advances in human factors and ergonomics in healthcare and medical devices. *AHFE 2021 Virtual Conference on Human Factors and Ergonomics in Healthcare and Medical Devices*. Springer Cham. https://doi.org/10.1007/978-3-030-80744-3

Khan, A. Q., Muhammad, S. G., Raza, A., & Pimanmas, A. (2025). Advanced machine learning techniques for predicting mechanical properties of eco-friendly self-compacting concrete. *Journal of Road Engineering*, *5*(2), 213–229. https://doi.org/10.1016/j.jreng.2024.12.002

Khumaidi, A., Purwanto, Y. A., Sukoco, H., & Wijaya, S. H. (2022). Using fuzzy logic to increase accuracy in mango maturity index classification: Approach for developing a portable near-infrared spectroscopy device. *Sensors*, *22*(24), 9704. https://doi.org/10.3390/s22249704

Krishnan, N., Ravi Kumar, K., & R., S. A. (2024). Solar radiation forecasting using gradient boosting based ensemble learning model for various climatic zones. *Sustainable Energy, Grids and Networks*, *38*, 101312. https://doi.org/10.1016/j.segan.2024.101312

Kumar, A., Pandey, A., Gupta, T., & Ghosh, S. K. (2025). Lube oil life prediction for heavy earth moving machinery (HEMM): A machine learning approach. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*. https://doi.org/10.1177/09544089241311377

Kurnianto, B. A., Prayogo, R. H., Nababan, N. P., & Suharjito. (2024). Predictive maintenance using linear regression. *Syntax Literate*, *9*(3). https://doi.org/10.36418/syntax-literate.v9i3

Liang, Z., Xie, F., Guo, Z., Wang, Z., Dou, H., Wang, B., & Shen, B. (2024). Optimization and prediction of a novel preignition in hydrogen direct injection engines through experimentation and the Random forest algorithms. *Energy Conversion and Management*, *313*, 118602. https://doi.org/10.1016/j.enconman.2024.118602

Liu, H., Yan, S., Huang, M., & Huang, Z. (2024). A fault diagnosis method for hydraulic system based on multi-branch neural networks. *Engineering Applications of Artificial Intelligence*, *137*, 109188. https://doi.org/10.1016/j.engappai.2024.109188

Ma, X., Tao, F., Zhang, M., Wang, T., & Zuo, Y. (2019). Digital twin enhanced human-machine interaction in product lifecycle. *Procedia CIRP*, *83*, 789–793. https://doi.org/10.1016/j.procir.2019.04.330

Meddeb, H., Abdellaoui, Z., & Ezzine, O. (2023). Human machine interface design of industrial automated machine using simatic scada system. *International Journal of Industrial Engineering*, *30*(2), 435–450. https://doi.org/10.23055/ijietap.2023.30.2.8561

Mekelburg, E., & Strauss, J. (2024). Pooling and winsorizing machine learning forecasts to predict stock returns with high-dimensional data. *Journal of Empirical Finance*, *79*, 101538. https://doi.org/10.1016/j.jempfin.2024.101538

Muhammed, D., Ahvar, E., Ahvar, S., Trocan, M., Montpetit, M.-J., & Ehsani, R. (2024). Artificial Intelligence of Things (AIoT) for smart agriculture: A review of architectures, technologies and solutions. *Journal of Network and Computer Applications*, *228*, 103905. https://doi.org/10.1016/j.jnca.2024.103905

Nguyen Thanh, P., & Cho, M.-Y. (2024). Advanced AIoT for failure classification of industrial diesel generators based hybrid deep learning CNN-BiLSTM algorithm. *Advanced Engineering Informatics*, *62*, 102644. https://doi.org/10.1016/j.aei.2024.102644

Peng, H., Chen, Y., Shangguan, L., Cheng, R., Li, Y., & Yang, C. (2023). Multi-objective-based intelligent lubrication system performance evaluation technology for construction machinery. *Applied Sciences*, *13*(21), 11768. https://doi.org/10.3390/app132111768

Perremans, Y., Peremans, B., Mrak, B., Schlimpert, S., & Tao, Z. (2024). The impact of advanced lubrication strategies on the performance of oil-lubricated multi-bearing systems. *Tribology International*, *194*, 109479. https://doi.org/10.1016/j.triboint.2024.109479

Probst, P., Wright, M. N., & Boulesteix, A. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, *9*(3), e1301. https://doi.org/10.1002/widm.1301

Rahman, M. H., Shahriar, S., & Menezes, P. L. (2023). Recent progress of machine learning algorithms for the oil and lubricant industry. *Lubricants*, *11*(7), 289. https://doi.org/10.3390/lubricants11070289

Rojas, L., Peña, Á., & Garcia, J. (2025). AI-driven predictive maintenance in mining: A systematic literature review on fault detection, digital twins, and intelligent asset management. *Applied Sciences*, *15*(6), 3337. https://doi.org/10.3390/app15063337

Sanap, P. V. C. (2025). Design and implementation of real time clock using RTC DS3231 and Arduino Uno. *International Journal for Research in Applied Science and Engineering Technology*, *13*(2), 1545–1551. https://doi.org/10.22214/ijraset.2025.67162

Sardhara, T., & Tamboli, K. (2018). Design and development of automatic lubrication system for ATC of CNC. *Materials Today: Proceedings*, *5*(2), 3959–3964. https://doi.org/10.1016/j.matpr.2017.11.653

Singh, G., Appadurai, J. P., Perumal, V., Kavita, K., Ch Anil Kumar, T., Prasad, D., Pazhani, A. A. J., & Umamaheswari, K. (2022). Machine learning-based modelling and predictive maintenance of turning operation under cooling/lubrication for manufacturing systems. *Advances in Materials Science and Engineering*, *2022*, 9289320. https://doi.org/10.1155/2022/9289320

Sinitò, D., Santarcangelo, V., Stanco, F., & Giacalone, M. (2023). Industry 4.0: Machinery integration with supply chain and logistics in compliance with Italian regulations. *MethodsX*, *11*, 102269. https://doi.org/10.1016/j.mex.2023.102269

Society of Tribologists and Lubrication Engineers. (2023). STLE Report on emerging issues and trends in tribology and lubrication engineering. https://test.stle.org/files/Publications/White_Papers/Emerging_Issues_and_Trends/files/White_Papers/2020_Emerging_Trends_Report.aspx?hkey=477ba040-218e-4c31-ae6a-6657dd246140

Surucu, O., Gadsden, S. A., & Yawney, J. (2023). Condition monitoring using machine learning: A review of theory, applications, and recent advances. *Expert Systems with Applications*, *221*, 119738. https://doi.org/10.1016/j.eswa.2023.119738

Ucar, A., Karakose, M., & Kırımça, N. (2024). Artificial intelligence for predictive maintenance applications: key components, trustworthiness, and future trends. *Applied Sciences*, *14*(2), 898. https://doi.org/10.3390/app14020898

Val, S., Lambán, M. P., Lucia, J., & Royo, J. (2024). Analysis and prediction of wear in interchangeable milling insert tools using artificial intelligence techniques. *Applied Sciences*, *14*(24), 11840. https://doi.org/10.3390/app142411840

Wang, X., Lü, H., Crow, W. T., Corzo, G., Zhu, Y., Su, J., Zheng, J., & Gou, Q. (2023). A reduced latency regional gap-filling method for SMAP using random forest regression. *IScience*, *26*(1), 105853. https://doi.org/10.1016/j.isci.2022.105853

Wszelaczyński, G., Capanidis, D., Paszkowski, M., & Leśniewski, T. (2021). Operating problems of lubrication of friction nodes in mining machines working in an aggressive environment. In J. Stryczek & U. Warzyńska (Eds), *Advances in Hydraulic and*

Pneumatic Drives and Control 2020 (pp. 228–238). Springer International Publishing. https://doi.org/10.1007/978-3-030-59509-8_20
Wu, F., Wang, Y., Zheng, M., Wang, J., Pan, J., & Liu, L. (2025). Prediction and quality zoning of potentially suitable areas for Panax notoginseng cultivation using MaxEnt and random forest algorithms in Yunnan Province, China. *Industrial Crops and Products*, *229*, 120960. https://doi.org/10.1016/j.indcrop.2025.120960
Yakoubi, S., Kobayashi, I., Uemura, K., Nakajima, M., Hiroko, I., & Neves, M. A. (2023). Recent advances in delivery systems optimization using machine learning approaches. *Chemical Engineering and Processing - Process Intensification*, *188*, 109352. https://doi.org/10.1016/j.cep.2023.109352
Yin, Y., Shi, P., Zhang, S., Qing, T., Zhou, N., Qian, L., & Chen, L. (2024). Lubrication failure mechanism of rolling-sliding steel ball against oil-impregnated porous polyimide and bearing steel in double friction pair. *Engineering Failure Analysis*, *164*, 108622. https://doi.org/10.1016/j.engfailanal.2024.108622
Zhang, Q., Han, X., Fang, X., Liu, M., Ge, K., & Jiang, H. (2024). Optical frequency multiplication using residual network with random forest regression. *Heliyon*, *10*(10), e30958. https://doi.org/10.1016/j.heliyon.2024.e30958
Zhang, W., Shi, P., Jia, P., & Zhou, X. (2024). A novel gradient boosting approach for imbalanced regression. *Neurocomputing*, *601*, 128091. https://doi.org/10.1016/j.neucom.2024.128091