KARYA ILHAM PUBLISHING
CONNECTING MINDS, BRIDGING IDEAS, INSPIRING READERS

International Journal of Advanced Research in Computational Thinking and Data Sciences

Journal homepage:
https://karyailham.com.my/index.php/ctds/index
ISSN: 3030-5225

Prediction of Tribological Behaviour using Machine Learning

Nur Aisyah Roslan[1], Shafie Kamaruddin[1,*], Mohd Hafis Sulaiman[2], Muhammad Hazman Sharuddin[1], Nor Aiman Sukindar[3], Ahmad Zahirani Ahmad Azhar[1], Mohamad Talhah Al Hafiz[1]

[1] Department of Manufacturing and Materials Engineering, Faculty of Engineering, International Islamic University Malaysia (IIUM), 53100 Jalan Gombak, Kuala Lumpur, Malaysia
[2] Department of Mechanical and Manufacturing Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[3] School of Design, Universiti Teknologi Brunei Jalan Tungku Link Gadong BE1410, Brunei Darussalam

ARTICLE INFO

Article history:
Received 28 July 2025
Received in revised form 15 August 2025
Accepted 18 September 2025
Available online 29 September 2025

Keywords:
Tribology; machine learning; random forest; decision tree; support vector regression

ABSTRACT

Tribology plays a pivotal role in determining the performance and durability of mechanical systems. Friction and wear reduce component life, increase costs, and affect system reliability. This study explores how Machine Learning (ML) models can predict tribological behavior-specifically coefficient of friction (COF), temperature, and worn area-based on experimental data from block-on-ring tests. Using lubricants and coatings as input parameters, three supervised ML algorithms were applied: Random Forest (RF), Decision Tree (DT), and Support Vector Regression (SVR). The dataset was pre-processed and split into training and testing sets. Hyperparameters were optimized using grid search. Results show RF provides the best accuracy for COF and worn area prediction, while DT performs best in predicting temperature. SVR showed the least accuracy across all outputs. These findings demonstrate the potential of ML as a predictive tool in tribology.

1. Introduction

The reliability and longevity of internal combustion engines are closely linked to the performance of their lubrication systems. Lubricants play a critical role in forming protective films on contact surfaces, thereby reducing friction, minimizing wear, and maintaining operational stability under varying loads and temperatures [1] . In recent years, there has been a growing shift in the automotive industry towards the use of low-viscosity oils, such as SAE 5W30 and SAE 0W20, primarily driven by the need to improve fuel efficiency and reduce emissions [2]. However, this transition presents new tribological challenges. Lower viscosity reduces fluid film thickness, potentially increasing direct contact between surfaces and accelerating wear mechanisms [3], especially under high-stress conditions.

* *Corresponding author.*
E-mail address: shafie@iium.edu.my

https://doi.org/10.37934/ctds.6.1.2837

To address these challenges, attention has turned toward enhancing the performance of lubricants through the incorporation of solid additives, particularly nanomaterials such as graphene [4] and fullerene [5], and the application of surface coatings [6]. These materials offer unique mechanical and thermal properties that can reduce friction and wear beyond the capabilities of conventional lubricant formulations [7,8]. Various experimental studies have reported improvements in tribological performance when such additives or coatings are used independently [9,10]. Nonetheless, the effects of their combined use, particularly under conditions representative of engine operation, remain insufficiently studied. Furthermore, a systematic evaluation of these advanced materials across multiple tribological indicators such as friction coefficient, wear area, and temperature rise is still lacking [4,11].

Experimental investigations in tribology are inherently complex, involving numerous interdependent variables including load, speed, temperature, material combinations, and lubrication regimes. Conducting comprehensive experimental trials is time-consuming and resource intensive. In this context, data-driven approaches, particularly those based on machine learning (ML), have emerged as a practical tool to analyze and predict tribological behavior [12]. ML algorithms can process large datasets to identify trends, model non-linear relationships, and provide accurate estimations of key performance indicators [13]. Such models have the potential to complement experimental methods by reducing the number of physical tests required and enabling more efficient material selection and lubricant design. Despite these advantages, the application of machine learning in tribological analysis especially in relation to engine lubricants modified with advanced additives and coatings is still in its early stages [13,14]. There remains a need for studies that not only develop predictive models but also validate them against empirical data obtained from standardized testing methods.

This research addresses this gap by developing and validating machine learning models to predict the coefficient of friction, worn area, and temperature using data from block-on-ring tribological tests involving various lubricants and coatings [15]. The significance of this work lies in its potential to improve understanding of tribological performance under realistic operating conditions, while also demonstrating how predictive modeling can be integrated into material evaluation processes. The objective is to provide a framework for optimizing lubricant and surface treatment combinations using a combination of experimental and computational methods.

2. Methodology

2.1 Data Collection

The dataset used in this study was sourced from previously published experimental work involving tribological tests performed using a block-on-ring configuration [15]. The experiments covered 10 types of lubricants, and 6 types of coatings applied to the ring surfaces. Parameters such as weight loss, contact surface temperature, angular speed, and torque were recorded at regular intervals of 60 seconds. These values were used to compute the coefficient of friction (COF), along with the worn area and temperature, which serve as the primary outputs of interest for this study (refer Table 1).

Table 1
Summary of experimental variables from [15]

Category	Details
Lubricants	10 types (e.g., SAE 5W30, SAE 0W20, etc.)
Coatings	6 types (e.g., DLC, TiN, CrN, etc.)
Measured at	60 seconds interval
Measured parameters	- Weight loss (mg) - Contact surface temperature (°C) - Angular speed (RPM) - Torque (Nm)
Derived outputs	- Coefficient of Friction (COF) - Worn Area (mm²) - Surface Temperature (°C)

2.2 Data Pre-processing

Prior to model development, the dataset underwent a series of pre-processing steps to ensure data consistency and compatibility with machine learning (ML) algorithms. Redundant or constant variables, such as material composition of the block and ring and applied load, which remained unchanged across all samples, were removed to prevent unnecessary model complexity. Categorical entries such as lubricant or coating names (e.g., “drilling oil”, “cutting fluid”) were encoded into numerical values to enable their interpretation by ML models.

Feature scaling was applied to normalize the numerical data and mitigate the influence of differing value ranges across features. Missing or outlier data points were reviewed and treated accordingly. To ensure random distribution and prevent bias, the dataset was also shuffled before further processing. As each ML model processes one target output at a time, the input features were prepared separately for predicting each of the three target outputs: COF, worn area, and temperature.

2.3 Data Splitting

The complete dataset was divided into training and testing subsets using the train-test split method. A ratio of 75:25 was applied, where 75% of the data was used to train the models and 25% was reserved for testing. The training dataset was used to establish the relationship between the input features and the target output, while the test dataset was used to evaluate the predictive accuracy of the trained models on unseen data.

2.4 Machine Learning Models

Three supervised learning algorithms were selected for model development: Random Forest (RF), Support Vector Regression (SVR), and Decision Tree (DT). These models were chosen based on their proven ability to handle nonlinear data, their robustness against overfitting (particularly in ensemble methods like RF), and their interpretability. Each model was trained and evaluated separately for the three target outputs. The comparative analysis aimed to determine which model performed best for each prediction task.

2.5 Hyperparameter Optimization

To improve the performance of the ML models, grid search optimization was employed to identify the most suitable hyperparameters. Grid search systematically evaluates all possible combinations of predefined hyperparameter values and selects the configuration that yields the best model performance on the training set. The optimized parameters were then used to retrain the models before final evaluation.

2.6 Model Evaluation

Model performance was assessed using a combination of standard regression metrics. The coefficient of determination (R^2) served as the primary metric, indicating how well the model explained the variance in the dataset. An R^2 value closer to 1 signifies better predictive performance. Acceptable model accuracy was defined as $R^2 \geq 0.70$, while values above 0.90 were considered excellent. In addition to R^2, error-based metrics were calculated, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics quantify the average deviation between predicted and actual values and provide further insight into model accuracy. Lower values across all three metrics indicate higher model precision. The equations used for these performance metrics are as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}| \tag{1}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y})^2 \tag{2}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y})^2} \tag{3}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2} \tag{4}$$

Where,
y_i - represents the actual values.
$\hat{y}$ - represents the predicted values.
$\bar{y}$- is the mean of the actual values.
N - is the number of data points.

3. Results and Discussion

Three ML models were used in this study; Random Forest (RF), Support Vector Regression (SVR), and Decision Tree (DT). The ML models were used to predict tribological behaviour which are COF, temperature and worn area. To evaluate the performance of the ML models, the *MSE, MAE, RMSE and R^2* for each model in predicting the COF, temperature and worn area of the block are shown in the Table 2 – Table 4.

3.1 Performance of Machine Learning (ML) models for Coefficients of Friction (COF)

Table 2 shows the performance of each ML models in predicting the COF for steel block. RF and DT models attained a decent value of R^2 which are 0.7828 and 0.6909. However, SVR model recorded a quite deviated value of R^2 which is -84.2046. The obtained negative result may be due to the inaccuracy of the experimental results and datasets. The Random Forest (RF) model achieved the highest R^2 value of 0.7828 for predicting the Coefficient of Friction (COF), indicating good performance. This model exhibits 78% accuracy in predicting COF for steel blocks and tribological test factors. The MSE, MAE, and RMSE values for this model were impressively low at 9.4970, 0.0022, and 0.0031, respectively, indicating high prediction accuracy.

Table 2
Summary of Performance of ML models for COF

ML models	MSE	MAE	RMSE	R^2
Random Forest	9.4970	0.0022	0.0031	0.7828
Support Vector Regression	0.0037	0.0560	0.0610	-84.2046
Decision Tree	1.3517	0.0023	0.0037	0.6909

Figure 1 - Figure 3 show comparison of actual and prediction of COF over observation number for RF, SVR and DT models respectively. Figure 1 shows a comparison between the actual and predicted coefficients of friction (COF) using the Random Forest model. Figure 2, corresponding to the SVR model, shows a considerable mismatch between the actual (blue dots) and predicted (red dots) values, with high fluctuations in the predicted values across observations. This indicates that the SVR model fails to capture the underlying pattern accurately, leading to poor prediction consistency. In contrast, in Figure 3, which represents the Decision Tree model, demonstrates a much closer alignment between the actual and predicted values. The predicted values (red dots) follow the actual trend (blue dots) with reduced deviations, suggesting that the Decision Tree model provides a more accurate and stable prediction of the COF compared to the SVR model.

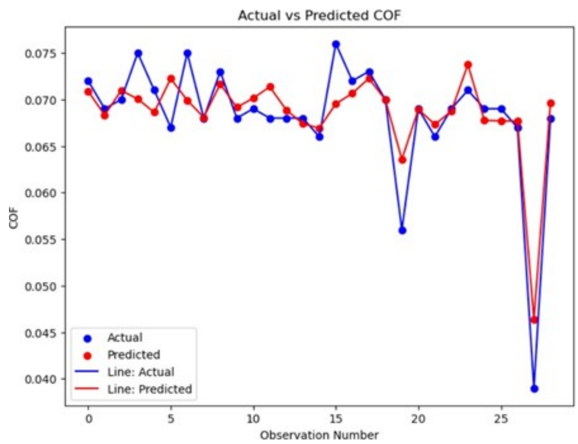

Fig. 1. Actual vs. Predicted value of COF by RF Model

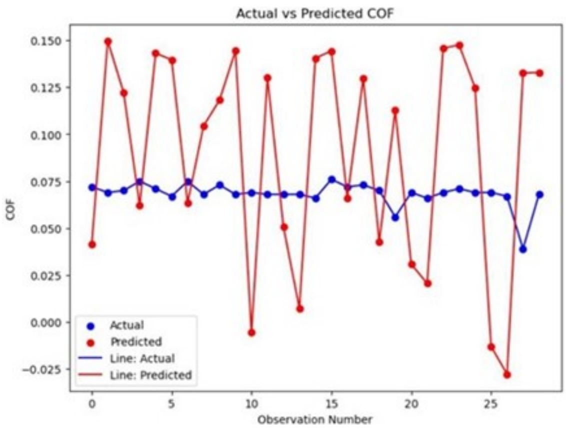

Fig. 2. Actual vs. Predicted value of COF by SVR Model

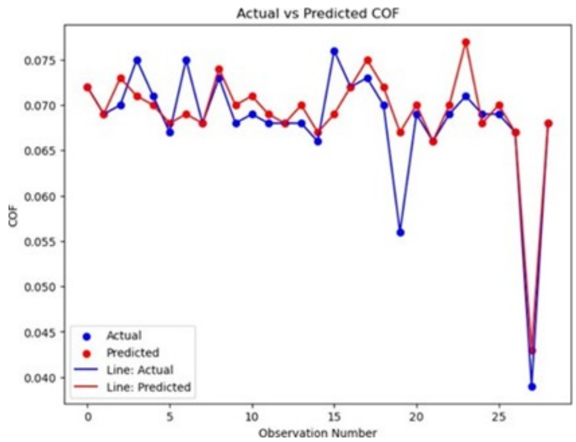

Fig. 3. Actual vs. Predicted value of COF by DT Model

3.2 Performance of Machine Learning (ML) models for Temperature

The performance indicators for machine learning (ML) models used to predict steel block temperatures are presented in Table 3. The test sets for the ML models exhibit R^2 values ranging from 0.1935 to 0.8250. In terms of Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), the ranges were 212.1415 to 977.6777, 9.0818 to 21.4694, and 14.5651 to 31.2678, respectively.

Table 3
Summary of performance of ML models for temperature

ML models	MSE	MAE	RMSE	R^2
Random Forest	338.0329	10.1399	18.3857	0.7212
Support Vector Regression	977.6777	21.4694	31.2678	0.1935
Decision Tree	212.1415	9.0818	14.5651	0.8250

Figure 4 compares the actual and predicted temperatures using the Random Forest (RF) model, showing a strong alignment between the two, indicating that RF effectively predicts temperature variations in the tribological system. The Decision Tree (Figure 6) also exhibits good predictive accuracy, with predicted temperatures closely matching actual measurements. However, Figure 5, representing the SVR model, reveals a significant discrepancy between the predicted and actual values, with larger fluctuations observed. This poor performance could be due to SVR's sensitivity to the choice of kernel function and hyperparameters, which may not be optimally tuned for this specific dataset.

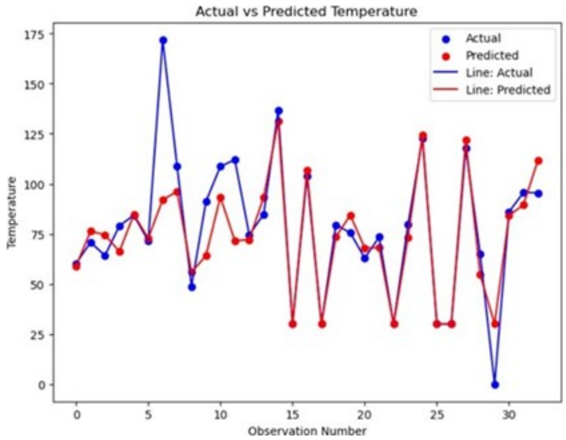

Fig. 4. Actual vs. Predicted value of lubricant temperature by RF Model

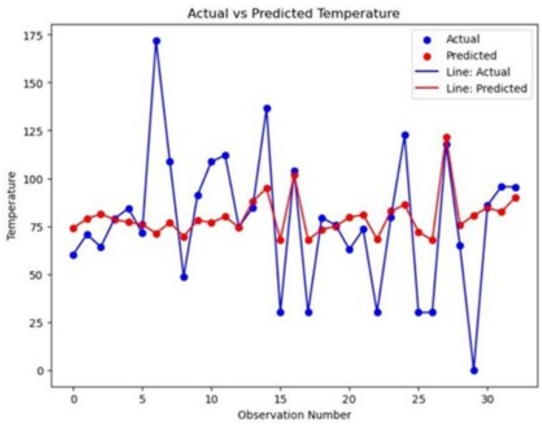

Fig. 5. Actual vs. Predicted value of lubricant temperature by SVR Model

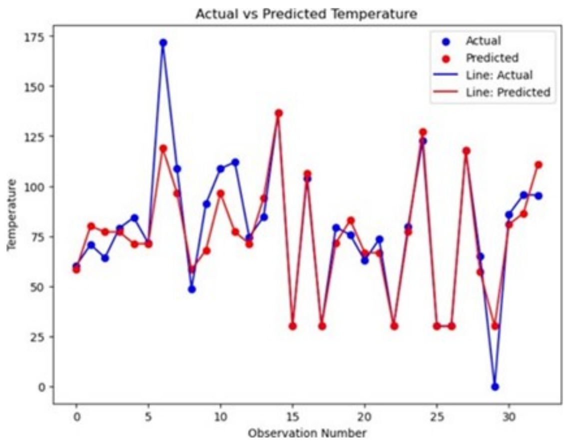

Fig. 6. Actual vs. Predicted value of lubricant temperature by DT Model

3.3 Performance of Machine Learning (ML) models for Worn Area

Table 4 shows Machine Learning (ML) model performance in predicting the Worn Area. The R^2 values on test sets ranged from 0.2797 to 0.8270 for the machine learning (ML) models. The Mean Squared Error (MSE) ranged between 54.0492 and 224.9788, Mean Absolute Error (MAE) ranged between 4.2374 and 9.5836, and Root Mean Squared Error (RMSE) varied between 7.3518 and 14.9993. Among the models, Random Forest (RF) recorded the highest R^2 value, which is 0.8270. This indicates that the RF model successfully identified the worn area of the block.

Table 4
Summary of Performance of ML models for Worn Area

ML models	MSE	MAE	RMSE	R^2
Random Forest	54.0492	4.7657	7.3518	0.8270
Support Vector Regression	224.9788	9.5836	14.9993	0.2797
Decision Tree	63.5641	4.2374	7.9727	0.7965

Figure 7, the actual versus predicted wear area values for the Random Forest (RF) model show a good fit, indicating that RF can predict wear characteristics accurately. Similarly, the Decision Tree model (Figure 9) also provides a reliable prediction with minimal deviation from actual values. However, the Support Vector Regression (SVR) model, as shown in Figure 8, struggles significantly, with a poor alignment between predicted and actual wear areas. The R-squared value of 0.2797 further supports this, suggesting a weak model fit. The reason for SVR's lower performance could be its limitations in handling non-linear relationships in the data, compared to RF and Decision Tree, which better capture the complexity of the wear patterns. The fluctuating predictions from SVR indicate it cannot generalize well for this dataset.

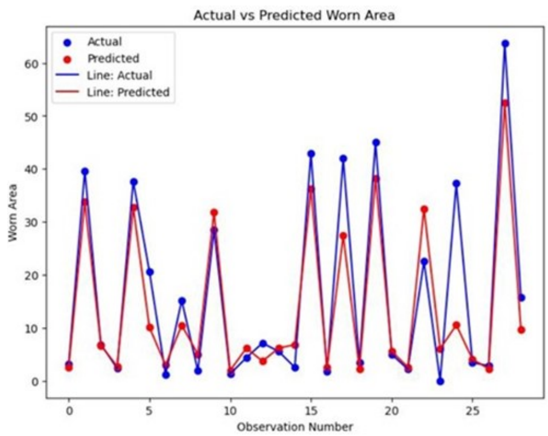

Fig. 7. Actual vs. Predicted value of worn area by RF Model

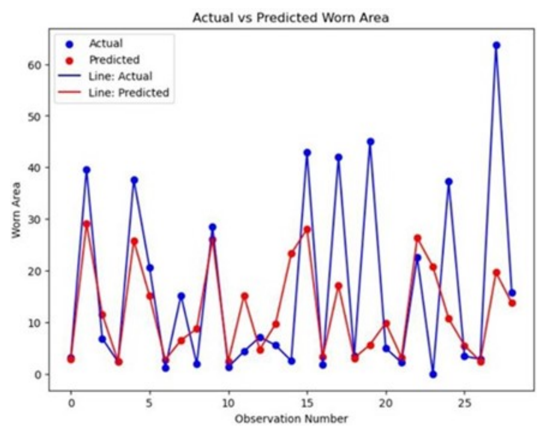

Fig. 8. Actual vs. Predicted value of worn area by SVR Model

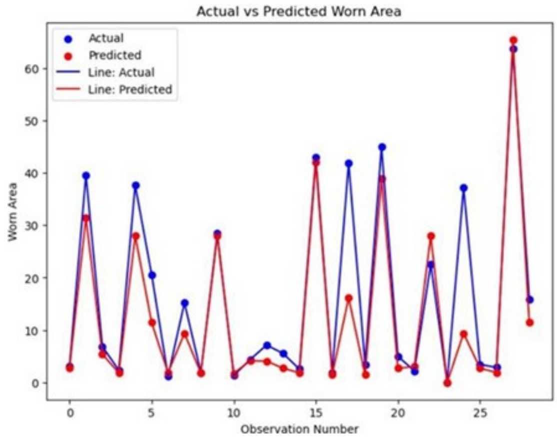

Fig. 9. Actual vs. Predicted value of worn area by DT Model

The main finding from this study highlights the predictive power of machine learning models, particularly Random Forest, in forecasting these tribological parameters. These findings emphasize the importance of selecting optimal lubricants and coatings for better performance in industrial applications, with machine learning offering valuable insights for optimization.

4. Conclusions

In conclusion, machine learning models were employed to predict and validate the coefficient of friction, worn area, and temperature. The Random Forest (RF) model performed best for predicting both the coefficient of friction and wear area, while the Decision Tree (DT) model excelled at predicting temperature. These findings demonstrate the potential of machine learning in enhancing the predictive accuracy of tribological behavior based on lubricant and coating properties.

Overall, this study contributes to optimizing lubrication strategies and improving predictive capabilities in tribology, offering a clearer path toward enhancing engine durability and performance through data-driven approaches.

Acknowledgement

This study was supported by RMC research grant (No. RMCG20-036-0036) awarded by the International Islamic University of Malaysia (IIUM) which made this study possible.

References

[1] Ideris, Muhammad Haziq, Shafie Kamaruddin, Mohd Hafis Sulaiman, Nor Aiman Sukindar, Ahmad Zahirani Ahmad Azhar, and A. S. H. M. Yasir. "Effects of Coating and Lubrication on Friction and Wear for Metal-to Metal Application." *J. Adv. Res. Appl. Mech* 110 (2023): 52-62. https://doi.org/10.37934/aram.110.1.5262

[2] Mo, Yimin, Jun Wang, Ye Hong, Xiao Yang, and Juncheng Lv. *Study on the Influence of Low-Viscosity Engine Oil on Engine Friction and Vehicle Worldwide Harmonized Light Vehicles Test Cycle Fuel Economy*. No. 2020-01-5062. SAE Technical Paper, 2020. https://doi.org/10.4271/2020-01-5062

[3] Zhang, Yanyan, Ziyuan Ma, Yan Feng, Ziyu Diao, and Zhentao Liu. "The effects of ultra-low viscosity engine oil on mechanical efficiency and fuel economy." *Energies* 14, no. 8 (2021): 2320. https://doi.org/10.3390/en14082320

[4] Marlinda, Ab Rahman, Gregory Soon How Thien, Mehmood Shahid, Ting Yang Ling, Abu Hashem, Kah-Yoong Chan, and Mohd Rafie Johan. "Graphene as a lubricant additive for reducing friction and wear in its liquid-based form." *Lubricants* 11, no. 1 (2023): 29.https://doi.org/10.3390/lubricants11010029.

[5] Rui, L. Ming, M. Hafis Sulaiman, M. Ashman Hakimi Azizan Ashman, M. Shuhaimi Ibrahim, M. Ridzuan Mohd Jamir, and Putra Malaysia. "Enhancing physical, mechanical, tribological and contact characteristics of water-based lubrication using ceramic nanomaterials MgO and SiC for metal-to-metal sliding interfaces." *Jurnal Tribologi* 39, no. 7 (2023): 51-68.

[6] Ibrahim, M. S., M. H. Sulaiman, R. Samin, Y. Yaakob, Shahira Liza Kamis, M. J. M. Ridzuan, A. A. Pauzi, N. A. Sukindar, and K. Bienk. "Tool wear of DLC coating as top-layered to CrN, TiAlSiN, TiAlN coatings in machining of steel and aluminum alloys." *Wear* 558 (2024): 205574. https://doi.org/10.1016/j.wear.2024.205574.

[7] Lu, Ziyan, Qingqing Lin, Zhaotao Cao, Wanyuan Li, Junjie Gong, Yan Wang, Kunhong Hu, and Xianguo Hu. "MoS2 nanomaterials as lubricant additives: A review." *Lubricants* 11, no. 12 (2023): 527. https://doi.org/10.3390/lubricants11120527

[8] Liu, Yanfei, Xiangyu Ge, and Jinjin Li. "Graphene lubrication." *Applied materials today* 20 (2020): 100662. https://doi.org/10.1016/j.apmt.2020.100662.

[9] Stojanović, Blaža, Aleksandar Vencl, Ilija Bobić, Slavica Miladinović, and Jasmina Skerlić. "Experimental optimisation of the tribological behaviour of Al/SiC/Gr hybrid composites based on Taguchi's method and artificial neural network." *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 40, no. 6 (2018): 311. https://doi.org/10.1007/s40430-018-1237-y.

[10] Sulaiman, M. H., R. N. Farahana, K. Bienk, C. V. Nielsen, and N. Bay. "Effects of DLC/TiAlN-coated die on friction and wear in sheet-metal forming under dry and oil-lubricated conditions: Experimental and numerical studies." *Wear* 438 (2019): 203040. https://doi.org/10.1016/j.wear.2019.203040.

[11] Guo, Yan-Bao, and Si-Wei Zhang. "The tribological properties of multi-layered graphene as additives of PAO2 oil in steel–steel contacts." *Lubricants* 4, no. 3 (2016): 30. https://doi.org/10.3390/lubricants4030030.

[12] Hasan, Md Syam, Amir Kordijazi, Pradeep K. Rohatgi, and Michael Nosonovsky. "Triboinformatic modeling of dry friction and wear of aluminum base alloys using machine learning algorithms." *Tribology International* 161 (2021): 107065. https://doi.org/10.1016/j.triboint.2021.107065.

[13] Marian, Max, and Stephan Tremmel. "Current trends and applications of machine learning in tribology—A review." *Lubricants* 9, no. 9 (2021): 86.. https://doi.org/10.3390/lubricants9090086.

[14] Jia, Binyuan, Qiang Wan, Long Yan, Yin Luo, Qiwen Wei, Chengqiang Niu, Bing Yang, Shanjun Li, and Liang Meng. "Tribological properties and machine learning prediction of FeCoCrNiAlN high entropy coatings." *Surface and Coatings Technology* 477 (2024): 130341. https://doi.org/10.1016/j.surfcoat.2023.130341.

[15] Sharuddin, Muhammad Hazman bin, M. H. Sulaiman, S. Kamaruddin, A. H. Dahnel, N. F. H. Abd Halim, M. J. M. Ridzuan, and Ahmad Majdi Abdul-Rani. "Properties and tribological evaluation of graphene and fullerene nanoparticles as additives in oil lubrication." *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology* 237, no. 8 (2023): 1647-1656.. https://doi.org/10.1177/13506501231175540