Vol. 46, No. 4 (2024) 624-638, DOI: 10.24874/ti.1736.08.24.09

Tribology in Industry

www.tribology.rs

RESEARCH

Engine Health Status Prediction Based on Oil Analysis With Augmented Machine Learning Algorithms

Slawomir Szrama[a]

[a]*Poznan University of Technology, Aviation Division, 60-965 Poznan, Piotrowo 3, Poland.*

Keywords:

Aircraft turbofan engine
Engine health status prediction
Artificial neural network
Turbofan engine lubrication system
Engine diagnostics and health monitoring
Engine oil sampling
Machine learning

* Corresponding author:

Slawomir Szrama
E-mail: slawomir.szrama@put.poznan.pl

Received: 22 August 2024
Revised: 20 September 2024
Accepted: 9 October 2024

A B S T R A C T

Tribology is the very efficient and strong tool in machine operations analysis. In the article author presented how the artificial intelligence algorithms could be applied to help in engine oil test results analysis. Based on the real-life turbofan engine oil sample test results dataset, the novel methodology of the machine learning algorithm implementation was presented. In order to take advantage of the artificial intelligence in engine oil test results interpretation, the augmented engine oil dataset was generated with additional predictors. Research case study was conducted for both original engine dataset as well as the augmented one. For the scientific purposed, various machine learning performance metrics were calculated, what allowed to precisely compare the results achieved for the original dataset and the one generated on the basis of the proposed novel method. The greatest achievement of the article was the presentation of the new methodology implementation in the real-life turbofan engine health status prediction. Presented methodology implemented into the aircraft (engine) maintenance management computer system allows to automate engine health status analysis and improve engine maintenance management.

© 2024 Published by Faculty of Engineering

1. INTRODUCTION

Modern aircraft turbofan engines are an extremely complex, sophisticated, and advanced machines. Their reliability, endurance as well as durability are the key factors affecting aircraft operations and flight safety. That is the reason why engine health status analysis and prediction is the crucial element in supporting flight operations and aircraft readiness. All engine components are operating under extremely challenging conditions while being affected by severe stresses, strains, and thermal loads. Even though modern turbofan engines are equipped with an advanced Full Authority Digital Engine Controllers (FADEC) which are able to monitor and control engine operations, still there are no embedded controllers to assess engine components wearing and degradation. The only element which allows aircraft mechanics to detect any premature engine degradation and abnormal wear are the magnetic chip detectors, which are designed to collect metal particles from the engine lubrication system. Unfortunately, they allow only to detect visible portion of the wearing product, while they are unable to sense any changes in chemical composition of the lubricants, which are the result of the unusual friction and thermal effects on the engine oil. That is the reason why tribology is such a strong tool in engine modules and parts wearing process analysis. During engine operations a slow, progressive wear metal concentration buildup above established abnormal criteria might be an indicator of impending engine or its component failure. Another indication could be a series of rapid wear metal concentration increases occurring below established abnormal criteria. What might be the typical sources of wear found in turbofan engines? Usually the worn bearings (balls, cages, races), bearing seals and retainers, bearing housing, constant speed drives, oil pump gears and gearbox castings. Nowadays, the analysis of the engine oil samples is being done manually by aviation laboratory personnel, which requires additional work to perform especially when the engine fleet is large and oil sample trending analysis is required. The question is, if it is possible to apply machine learning algorithms in engine premature degradation detection and prediction. In the article, a novel approach with full process methodology was presented on how to support maintenance personnel in engine health status prediction.

2. RESEARCH GAP AND MOTIVATION

Contemporary aircraft engines are usually maintained in accordance with the condition-based maintenance strategy. It means that we can continue engine flight operations as long as its condition, health and performance comply with the designed requirements and do not exceed the assumed range. What was the motivation of the study? It results out of the experience in the air force and aviation industry. One of the most crucial decisions maintenance managerial personnel must make is answering the question: Can we still continue engine flight operations, or should we remove engine from the aircraft, before it fails. Too early engine removal decreases efficiency of engine useful life usage and is not financially effective. It also results in mission capability rate degradation and reduces aircraft fleet operations. On the other hand, continuing engine flight operations while engine should be removed from service might result in aircraft accident or even catastrophe. That is why it is extremely crucial to determine engine health status and predict the moment when it is absolutely necessary to stop engine operations. Proposed methodology helps airline operators in the planning process of the engine replacement, repairs and overhauls, which are extremely expensive and usually require spare engines which are not available at any moment.

3. LITERATURE REVIEW

Machine Learning (ML) is a branch of artificial intelligence that studies algorithms able to learn autonomously, directly from the input data. ML techniques have made a huge leap forward. That is why researchers have started to consider ML for applications within the industrial field. ML was considered to be applied in maintenance and quality management, with failure modes classification and prediction, condition monitoring and fault detection or downtime minimization and maintenance planning.

However, industrial applications are still few and limited. What are the emerging areas of machine learning applications in industry. For example, Feng et. al. [1] proposed a transfer learning algorithm for gear wear severity assessment. In 2023 Feng et. al [2] analysed vibration-based gear wear monitoring and prediction techniques. With machine learning algorithms and neural network models, continuous values can be predicted and individual groups can be classified. Machine learning and neural networks could also be applied to the analysis of research results in a broad context. Using machine learning, it is possible to effectively determine correlations between individual tribological parameters and their strength, classify the test samples, and determine their influence on individual parameters such as wear area or depth, like presented by Marian and Tremmel [3], Rosenkranz et.al [4] or Hasan et.al [5].

Even though machine learning algorithms have been applied in some aspects of the software engineering, still not many publications could be found in the area of the tribology. There are some articles concerning very specific elements and components of the machines. The latest information on how the artificial intelligence with the machine learning algorithm could be applied to tribology was presented by Paturi et.al [6]. Authors conducted thorough examination of the role of machine learning in tribological research and concluded that by combining machine learning methods with tribological experimental data, interdisciplinary research could be conducted to better understand efficient resource utilization and resource conservation. Rahman et.al [7] in the complex review provided a thorough examination of seven machine learning algorithms applied in the field of tribology. Researchers took advantage of artificial neural networks, support vector machines, and decision trees, which have been successfully applied to predict lubricant properties, such as viscosity, coefficient of friction COF, and wear, under different operating conditions. Another review of recent advances and applications of machine learning in tribology were presented by Sose et.al [8], where they demonstrated successful implementation of neural networks, supervised, and stochastic learning approaches in identifying structure-property relationships. Rosenkranz et.al [9] discussed perspectives and presented some of the advances achieved by artificial intelligence implementation, specifically artificial neutral networks, towards tribological research. Another review of the machine learning applications into the tribology areas was presented by Marian and Tremmel [10] or in [11] where they proposed physics-informed machine learning algorithm implementation. A physics-informed neural network (PINN) was also proposed by Chen et.al [12] for fatigue life prediction with small amount of experimental data enhanced by physical models describing the fatigue behaviour of materials, while Deng et.al [13] and Rom M. [14] proposed this type of network to solve the Reynolds equation for finite journal bearings. Ni et.al [15] applied modified Physics-Informed Residual Network (PIResNet) for rolling element bearing fault diagnostics. Zhao et.al [16] applied the physics-informed neural network (PINN) to the hydrodynamic lubrication analysis. The same type of neural network was used by Cheng et.al [17] for hydrodynamic lubrication with cavitation. The other publications related to AI applications in tribology were presented by Hasan and Nosonovsky [18] who reviewed ML algorithms used to establish correlations between the structures of metallic alloys and composite materials, tribological test conditions, friction and wear. Another area of AI which was applied to tribology were the neural networks. For instance, Walker et.al [19] proposed application of the neural network to improve computational efficiency of tribo-dynamic simulations of machine elements without comprising accuracy relative to the numerical solution. Another neural network application was presented by Wang and Tsai [20] generated lubrication model was used to instantly predict thermohydrodynamic lubrication performance with adequate accuracy. Since tribology is a very powerful tool in machines health status prediction, that is the reason why there were some ideas of machine learning applications for machine parts and elements health status prediction. For instance, ML algorithms were tested and successfully applied by Hasan et.al [21] for friction and wear prediction of Al-graphite composites. Applications of the artificial intelligence in rotary machines fault detection were proposed by Liu et.al [22]. Zhao et.al [23] analysed deep learning neural networks in machine health monitoring.

Nevertheless, there are no publications in which augmented machine learning algorithms were applied to predict real-life turbofan engine health status on the basis of the original and real oil sampling data.

4. RESEARCH OBJECT DESCRIPTION

The research object selected for the case study and application of the machine learning algorithms in health status prediction was the turbofan engine equipped with the very advanced lubrication system. This power plant was a low bypass, high compression ratio, dual spool, turbofan engine incorporating a mixed flow augmentor. The engine is based on the modular concept allowing functionally and physically associated parts to be removed as modules. These modules are treated as individual entities rather than as a section of an engine. This concept increases maintainability since an engine can be returned to service more rapidly by installing a serviceable module rather than waiting for its own damaged module to be repaired. The five modules that make up the engine are as follows: inlet/fan module, core engine module, fan drive turbine module, augmentor duct and nozzle module, and gearbox module. The main area of interest was focused on the engine lubrication system. The purpose of the turbofan engine lubrication system is to provide oil to all engine bearing compartments, bearings and gears of the engine accessory drive gearbox, in order to lubricate, clean and cool all the components. Gears, located in the gearbox transmit torque between the gearbox and the engine accessories, rear compressor and the aircraft gearbox. All bearings and gears experience friction during engine operation so lubricating oil is supplied to the bearings and gearbox. The oil removes excess heat from the bearing compartments and gearbox to help extend the life of the component. The heat is removed through the fuel/oil cooler. The lubrication system (Figure 1) is divided into three subsystems. Pressure Subsystem which supplies clean, cool oil to the bearings and gears for lubrication and cooling. Scavenge Subsystem which collects oil from the bearing compartments and gearbox and returns it to the oil tank. Breather Subsystem which maintains proper air pressure in the bearing compartments to prevent oil leakage into the engine gas-path and prevent scavenge pump cavitation.

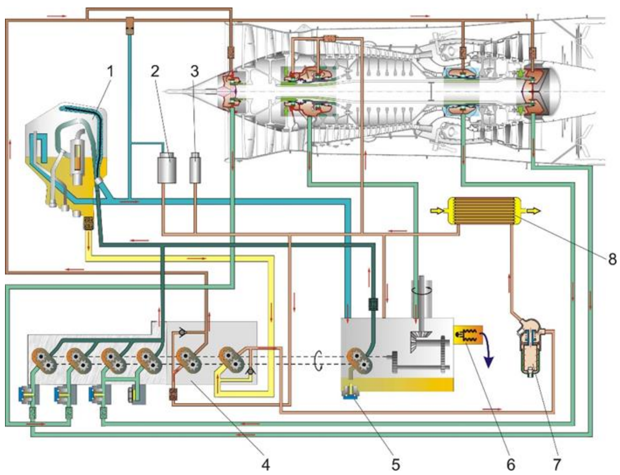

Fig. 1. Turbofan engine lubrication system where: 1 – oil tank, 2 – main oil pressure transmitter, 3 – low oil pressure transmitter, 4 – main oil pump with 3 chip detectors, 5 – chip detector No. 2 &3 bearings, 6 – breather pressurizing valve, 7 – oil filter, 8 – fuel-oil cooler.

One of the strongest tools in engine health status determination and prediction is the Oil Analysis Program (OAP). This program is a maintenance diagnostic tool which on the basis of the oil samples allows to monitor engine health by detecting the presence of metal particles (wear metals) in the oil. The OAP process provides early warning of a pending oil problem and ensures the integrity of the engine. Oil samples are usually performed with the very specific intervals. For the turbofan engine being the research object, the requirement is to sample oil after every flight operation (during combined postflight/preflight inspection). Oil sampling must be performed before within 30 minutes after engine shutdown and before the engine oil is serviced to prevent dilution of the sample.

The whole program is based on the spectrometric oil analysis, which is aimed to determine the type of amount of wear metals in lubricating fluid samples. The presence of unusual concentrations of an element in the oil sample can indicate abnormal wear of the engine. It allows to detect and prevent any engine failures.

5. RESEARCH METHODOLOGY

Research case study was based on spectrometric oil analysis, which is a diagnostic maintenance tool used to determine the type and amount of wear metals in lubricating fluid samples. Engines, transmissions, gearboxes and hydraulic systems are the types of equipment most frequently monitored. The presence of unusual concentrations of the elements in the fluid sample could indicate abnormal wear of the engine and/or its components. When the abnormal wear is verified, engine may be removed from service before a major failure of a fluid wetted component occurs. Spectrometric oil analysis enhances aircraft flight safety and engine readiness and serves as a decisive, preventive maintenance tool. Engine oils sample test results may be used as guidelines to assist in identifying incipient mechanical failures or in determining the quality and useful life of the engine oil. That is how the potential engine wear of failure and premature lubricant failure may be detected prior to a major engine failure or an expensive repairs/overhauls. Oil analysis may also be used to identify inadequate or improper maintenance procedures and low standard engine components or its parts.

During engine operations wear metals are generated by friction between moving metallic surfaces in mechanical part of the engine modules and components. Wear metal generation occurs also in all wetted subsystems, bearing compartments and gearboxes, to some degree and the lubricant serves as a repository for the wear metals. Metal elements concentration may also be generated from corrosive action resulting from moisture and electrolytic action within lubricated areas. We may conclude that the information related directly to the engine health status exists in the circulating lubricating fluid. This conclusion is developed as follow: first, the metal particles rubbed or gouged off the metal alloy surfaces will always have the same chemical compositions as the alloys from which they came. Second, the normal level and rate of production of each kind of metal particle can be established through oil analysis over a period of time. When an abnormal level and/or rate of production of wear metals is detected, the chemical identity of the abnormally produced particles will provide clues concerning the identity of the engine parts being worn.

In figure 2 a theoretical plot of wear concentration in parts per million (ppm) vs. engine operating hours or engine cycles was presented. Any condition which alters the normal relationship or increases the normal friction between moving parts will generally accelerate the rate of wear and increase the quantity of wear metal particles produced. An abnormal condition of this type will rapidly increase the concentration and rate of buildup of wear metals in stable fluid systems. If the condition is not detected and corrected, the deterioration will continue to accelerate, usually with major secondary damage to other parts of the engine or its assembly, resulting in the aircraft accident.

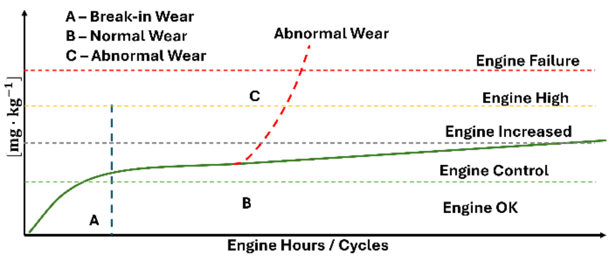

Fig. 2. Wear metal concentrations vs. Engine Hours/Cycles.

Wear metals produced in fluid lubricated mechanical assemblies can be measured in extremely low concentrations by spectrometric analysis of the oil samples taken from the engine. In the research case study, it was the atomic emission rotrode spectrometer (fig. 3) used to analyse collected engine oil samples. An emission spectrometer is an optical instrument used to determine the concentration of wear metals in lubricating fluid. The analysis is accomplished by subjecting the oil sample to a high voltage spark or plasma which energizes the atomic structure of the metallic elements, causing the emission of light. The intensity of the emitted light is proportional to the quantity of the element present in the sample allowing the concentration of that element to be determined. The light has a specific frequency or wavelength determined by the energy of the electron in transition.

Fig. 3. Rotating Disc Electrode Optical Emission Spectrometer.

Engine oil sampling procedure was introduced into the engine maintenance procedures. Oil sample was taken from the engine oil tank sample port (fig. 4).

Fig. 4. Engine oil tank sample port.

Engine oil sample was taken after every flight and must be taken prior to servicing the tank to prevent dilution of the sample. Every sample was taken within 30 minutes after engine shutdown. If the sample could not be obtained within 30 minutes after engine shutdown, engine was restarted and ran at idle for 10 minutes to obtain adequate oil sampling.

Engine oil sample elements concentrations dataset. The whole dataset consisted of the collections of the test results of 53 turbofan engines operating for the last 10 years. It resulted in 10371 records. Each record had 18 features: engine serial number, aircraft number, data of test, and 15 elements: Fe, Ag, Al, Cr, Cu, Mg, Na, Ni, Pb, Si, Sn, Ti, B, Mo, Zn. The whole engine test results dataset was divided into two sets: training and testing dataset consisting of the 7174 and 3197 records respectively. All engines were manufactured within the same 2-year period. However, they had various numbers of the engine flight hours / cycles collected. Nevertheless, all of the engines have not reached first scheduled overhaul. This fact is especially important, as the complete overhaul might change oil analysis result throughout the whole engine life cycle.

One of the crucial steps in engine health status prediction was to design Engine Health Status Model. Engine Health Status Model (EHStatus Mdl) was constructed on the basis of the OAP Atomic Emission Rotrode Limits and wear concentration levels and presented in table 1.

What is worth noticing is the fact that Engine Health Status was determined not only on the basis of the elements' concentration level, but also on the basis of the abnormal trends tracked for the last ten engine flights.

Table 1. Engine Health Status Model architecture.

EHStatus	OAP Atomic Emission Rotrode Limits [$mg \cdot kg^{-1}$]										
	Ag	**Al**	**B**	**Cr**	**Fe**	**Ni**	**Si**	**Ti**	**Zn**	Level	Engine Health Index Number
Engine Control	2	4		2	4	2		4		Abnormal Trend (ppm Increase in 10 Hrs)	1
Engine OK	0÷2	0÷10		0÷4	0÷10	0÷4		0÷10		Normal Range	5
Increased	N/A	11÷12		5	11÷12	5		11÷12		Marginal Range	4
High	3	13÷14		6	13÷14	6		13÷14		High Range	3
Failure	≥ 4	≥ 15	≥ 10	≥ 7	≥ 15	≥ 7	≥ 15	≥ 15	≥ 8	Abnormal	2

Engine training dataset was trained with the various machine learning algorithms like ensemble learner, Naive bayes, Support Vector Machines SVM, and tree based. Total number of iterations for the training was set at 105. Maximum time for training and validation was not set. Training was stopped when the loss function reached its minimum values or when the gradient of change for the loss function increased for the six following trials.

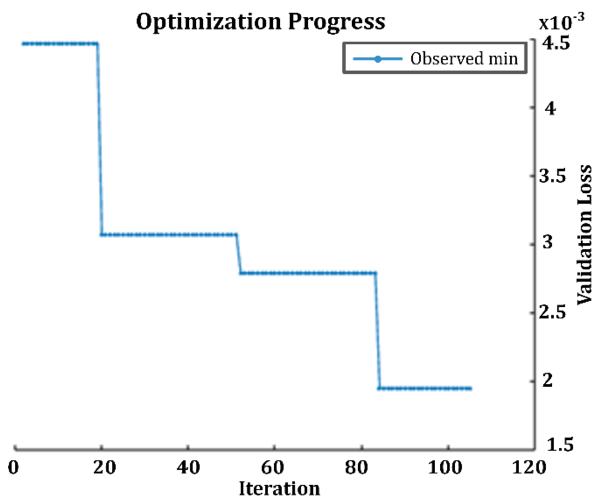

Fig. 5. Optimization Progress for the training process with various hyperparameters.

In order to optimize the training algorithm, the whole training process was optimized for all the machine learning algorithms and various hyperparameters.

In figure 5 training optimization process was presented. During this process the validation loss is calculated in each following iteration.

The main purpose of this step was to determine and select the best training machine learning algorithm for engine health status prediction. The best training results were achieved for Naive Bayes model with 0.0019515 observed validation loss and kernel distribution with width equal to 0.0649. The naive Bayes classifier is designed for use when predictors are independent of one another within each class, but it also works well in practice even when that independence assumption is not valid. It classifies data in two steps: training step, in which on the basis of the training data, the method estimates the parameters of a probability distribution, assuming predictors are conditionally independent given the class and prediction step, where for any unseen test data, the method computes the posterior probability of that sample belonging to each class. The method then classifies the test data according to the largest posterior probability. The class-conditional independence assumption greatly simplifies the training step since it is possible to estimate the one-dimensional class-conditional density for each predictor individually. While the class-conditional independence between predictors is not true in general, research shows that this optimistic assumption works well in practice. This assumption of class-conditional independence of the predictors allows the naive Bayes classifier to estimate the parameters required for accurate classification while using less training data than many other classifiers. This makes it particularly effective for data sets containing many predictors. The training step in naive Bayes classification is based on estimating P(X|Y), the probability or probability density of predictors X for the given class Y. The kernel distribution applied in the article is appropriate for predictors that have a continuous distribution. It does not require a strong assumption such as a normal distribution and it can be used in cases where the distribution of a predictor may be skewed or have multiple peaks or modes. It requires more computing time and more memory than the normal distribution. For each predictor, the naive Bayes classifier computes a separate kernel density estimate for each class based on the training data for that class. By default, the kernel is the normal kernel, and the classifier selects a width automatically for each class and predictor.

As a following step the best training algorithm was implemented and applied to the engine test dataset. In order to analyse engine health status predictions there were some performance metrics introduced.

One of the most important neural network performance metrics is the cross-entropy loss function (PRF). The function returns a result that heavily penalizes outputs that are extremely inaccurate ($\tilde{y}_i \sim 1-y_i$), with very little penalty for fairly correct classifications ($\tilde{y}_i \sim y_i$). Minimizing cross-entropy allow to converge to classification model.

where:

y_i – the following target value,

$\tilde{y}_i$ – the following predicted value,

Cross-entropy (PRF) could be calculated in accordance with equation 1.

$$CrossEntropy = PRF(\tilde{y}, y) = -\sum_{i=1}^{N} y_i ln\tilde{y}_i \quad (1)$$

In addition to the previously mentioned performance metrics, some more were added as a part of the comparison between the results achieved for the simulated and real-life engine data.

Accuracy performance metric. This performance metric is usually used in the case of the classification neural networks, where it is calculated as a ratio of the sum of the True Positive and True Negative values divided by the sum of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). It could be calculated in accordance with equation 2.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

The next performance metric is the mean absolute percentage error MAPE which is calculated between the predicted values and the actual values. It can be defined as:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i - \tilde{y}_i}{y_i} \cdot 100\% \quad (3)$$

Another performance metric, which is used in the Recurrent Neural Networks (RNN) is the coefficient of determination R-squared. In the context of regression, it is a statistical measure of how well the regression line approximates the actual data. R-squared coefficient can be calculated in accordance with equation 4.

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^{N}(y_i - \tilde{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (4)$$

where:

N – number of observations,

SSR – sum squared regression is the sum of the residuals[1] squared,

SST – total sum of squares is the sum of the distance the data is away from the mean all squared,

yi – the following target value,

$\tilde{y}_i$ – the following predicted value,

$\bar{y}$ – the mean of the predicted value

[1] residual value = actual y value – predicted y value

Percentage of errors %Error which the sum of the mispredictions divided by the number of elements in the dataset.

$$\%Error = \sum_{i=1}^{N} \frac{(y_i \sim= \tilde{y}_i)}{N} \cdot 100\% \quad (5)$$

Another comparison of the results might be performed by comparing relative accuracy RA, which could be calculated as a ratio of predictions to the actual values:

$$RA = \frac{\tilde{y}_i}{y_i} \cdot 100\% \quad (6)$$

Precision is the metric which presents how accurate the positive predictions are, and could be calculated as follows:

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

Recall is a metric that measures how often a machine learning model correctly identifies positive instances (true positives) from all the actual positive samples in the dataset and is calculated as follows:

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

The F1 score is calculated as the harmonic mean of the precision and recall scores, as shown below. It ranges from 0-100%, and a higher F1 score denotes a better-quality classifier.

$$F1Score = \frac{2 \cdot Precision \cdot Recall}{Precision+Recall} \quad (9)$$

One of the new machine learning performance metrics introduced into the case studies is the: Matthew's correlation coefficient (MCC). MCC is the best single-value classification metric which helps to summarize the confusion matrix or an error matrix. A confusion matrix has four entities: True positives (TP), True negatives (TN), False positives (FP) and False negatives (FN). Contrary to other well-known performance metrics like: F1score or prediction accuracy, MCC is related to all four confusion matrix entities. The formula to calculate MCC is presented in equation 10.

$$MCC = \frac{TN \cdot TP - FN \cdot FP}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (10)$$

MCC metric reflexes how good are the prediction rates for all four of these entities. MCC is said to be a reliable measure producing high scores. In order to suit most correlation coefficients, MCC also ranges between +1 and -1, where: +1 is the best agreement between the predicted and actual values and 0 is no agreement (meaning, prediction is random according to the real actual values). MCC helps to identify the ineffectiveness of the classifier in classifying especially the negative class samples.

What are the differences between MCC and F1 score? To evaluate binary classifications and their confusion matrices, scientists take advantage of several statistical rates, accordingly to the goal of the experiment they are investigating. Despite being a crucial issue in machine learning, no widespread consensus has been reached on a unified elective chosen measure yet. Accuracy and F1 score computed on confusion matrices have been (and still are) among the most popular adopted metrics in binary classification tasks. However, these statistical measures can dangerously show overoptimistic inflated results, especially on imbalanced datasets.

The Matthews correlation coefficient (MCC), instead, is a more reliable statistical rate that produces a high score only if the prediction obtained good results in all of the four confusion matrix categories (true positives, false negatives, true negatives, and false positives), proportionally both to the size of positive elements and the size of negative elements in the dataset.

6. RESULTS

Confusion matrix for the original model and engine dataset was presented in figure 6.

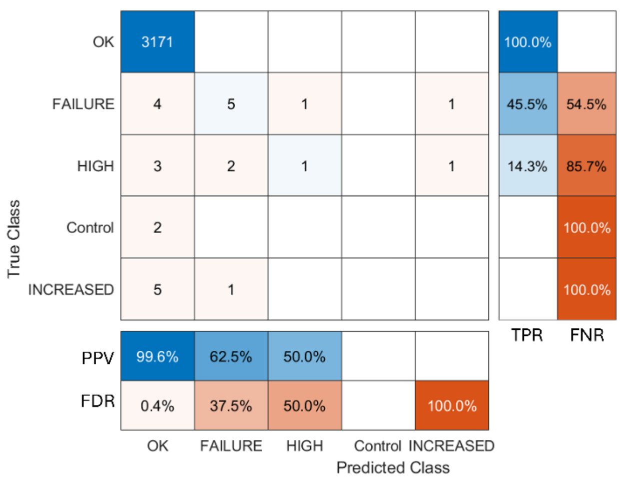

Fig. 6. Confusion matrix for the original engine dataset.

Confusion matrix was sorted according to the positive predictive value PPV, and the cell values were normalized across each column. As it might be deduced from the confusion matrix, True Positive Rate TPR for good engine with engine health status OK was 100%, for engine with Failure it was 45.5% and for engine with High concentration of the metal elements it was 14.3%. False Negative Rate FNR for engine Control and Increased Status was 100%, which means that there were no correct engine health status predictions. The Positive Predicted Values PPV was 99.6% for OK engine health status, 62.5% for Failure and 50% for High.

As it might be easily noticed from the confusion matrix the results achieved for the original dataset and engine health status model were not satisfying. In order to improve the machine learning performance, it was decided to implement very advanced automated features engineering function.

In addition to the original 15 predictors, new 35 features were generated to check the machine learning algorithm performance and accuracy. Before passing original engine training data to a classifier, new additional features were generated from the predictors in the engine dataset. The returned data was used to train the classifier. As a part of the predictors generation, the minimum redundancy maximum relevance (MRMR) features selection method was implemented. The MRMR algorithm finds an optimal set of features that is mutually and maximally dissimilar and can represent the response variable effectively. The algorithm minimizes the redundancy of a feature set and maximizes the relevance of a feature set to the response variable. The algorithm quantifies the redundancy and relevance using the mutual information of variables-pairwise mutual information of features and mutual information of a feature and the response. As a part of the MRMR feature selection, predictors were ranked, and then included into the requested number of top-ranked features in new training engine dataset.

In order to better understand the generated features, it was decided to generate information about the transformation function. The results were presented in table 2. Some generated features are a combination of multiple transformations. For example, some features were generated by converting the variable to a categorical variable with k-means clustering. K-means clustering is a type of unsupervised learning. The main goal of this algorithm to find groups in data and the number of groups is represented by K. It is an iterative procedure where each data point is assigned to one of the K groups based on feature similarity. K-means algorithm starts with initial estimates of K centroids, which are randomly selected from the dataset. The algorithm iterates between two steps assigning data points and updating centroids.

Table 2. Results of the transformation function applied to original dataset.

Feature Name	Data type	Input Variables	Transformations
kmi	Categorical	all valid numeric variables	Cluster index encoding (kmeans clustering with k = 10)
Al.*Ti	Numeric	Al, Ti	Al .* Ti
kmd10	Numeric	all valid numeric variables	Euclidean distance to centroid 10 (kmeans clustering with k = 10)
Ni+Ti	Numeric	Ni, Ti	Ni + Ti
kmd4	Numeric	all valid numeric variables	Euclidean distance to centroid 4 (kmeans clustering with k = 10)
Sn-Ti	Numeric	Sn, Ti	Sn - Ti
kmd8	Numeric	all valid numeric variables	Euclidean distance to centroid 8 (kmeans clustering with k = 10)
kmd3	Numeric	all valid numeric variables	Euclidean distance to centroid 3 (kmeans clustering with k = 10)
kmd9	Numeric	all valid numeric variables	Euclidean distance to centroid 9 (kmeans clustering with k = 10)
Cr+Ti	Numeric	Cr, Ti	Cr + Ti
kmd6	Numeric	all valid numeric variables	Euclidean distance to centroid 6 (kmeans clustering with k = 10)
Fe.*Ti	Numeric	Fe, Ti	Fe .* Ti
kmd1	Numeric	all valid numeric variables	Euclidean distance to centroid 1 (kmeans clustering with k = 10)

The automated features were added to the original data in order to enhance the machine learning performance and engine health status simulation. On the basis of the newly generated engine dataset, the original test engine dataset was transformed into an enhanced collection of records with 50 predictors. Confusion matrix for the original model and engine dataset was presented in figure 7. As it might be easily deduced from the confusion matrix presented in figure 7 for the augmented engine dataset the results are definitely better. In this case scenario True Positives Rate for Engine Health Statuses: Control, Failure and OK equalled 100%, while for High and Increased it was 57% and 50% respectively. The results achieved with the novel algorithm are especially important due to the fact that Engine Health Status: Failure was predicted in 100% cases. This situation is extremely important for the engine flight safety.

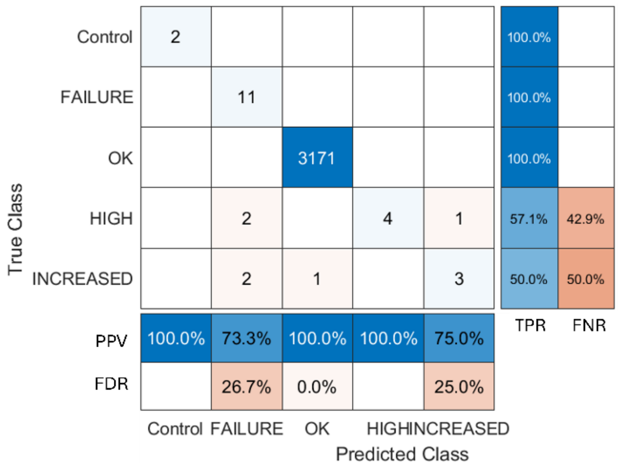

Fig. 7. Confusion matrix for the augmented engine dataset with automated features engineering.

Even very minor misprediction rate might result in engine failure and aircraft serious incident or accident. The Positive Predicted Values PPV was 100% for OK, Control and High engine health statuses, 73.3% for Failure and 75% for Increased Engine Health Status.

The results achieved for all the performance metrics were presented in Table 3.

Comparing the results of the machine learning algorithms achieved and presented in table 3 it is worth mentioning that considering performance PRF, accuracy, precision, recall, F1score and relative accuracy RA, the higher the results achieved are the better machine learning algorithms are working, and engine health status predictions are more accurate.

As far as the MAPE, MSE, RMSE, R2 coefficient and MCC are concerned, the lower the result is the better the predictions are. For these performance metrics, results which are closer to zero values represent the situation when the misprediction rate is very low, and all the predictions are close to the target values. As it might be easily deduced from the results presented in table 3, we may conclude that the proposed novel methodology resulted in the increment in the prediction error minimalization ranged from 34.5% for MCC to 87.5% for MAPE. For MSE achieved improvement equaled 85.5%, RMSE 62% and R2 54.4%.

Table 3. Machine Learning algorithm performance results for original and augmented model.

Model	PRF	Accuracy	Precision	Recall	F1Score	RA	MAPE	MSE	RMSE	R2	MCC
Original Model Dataset	0.9962	0.9974	0.9949	0.9999	0.9974	20.2365	0.5656	0.0262	0.1619	0.4266	0.6146
Augmented Model Dataset	0.9990	0.9992	0.9996	0.9999	0.9998	20.1918	0.0703	0.0037	0.0612	0.9353	0.9392

In figure 8 prediction results were presented for one of the engines with the original engine model and dataset. For engine #049 engine health status began to decrease after 4th oil sample testing and changed to the Increased level. Then it stayed for the next three flights at the same level when it decreased again to level 3 which is High Level of concentration. Again, the number of particles stayed at the same level for the following 3-4 flights and decreased again to the abnormal level and health index 2, being the engine Failure level. This is the situation where it is not safe to continue engine flight operations, as it might result in engine in-flight failure and aircraft accident. Unfortunately, engine health status prediction performed by the machine learning algorithm was not accurate and the prediction did not reflect the real engine health status. Even though it decreased after similar number of flights, but to the incorrect level being Failure not the High. What is worse is that after a few flights it returned to OK engine health status, while in real the engine kept degrading. Analyzing all the engines it is easy to notice that for most of them, engine health status prediction was close to the target values. Only for two engines there are some discrepancies, where predictions differ from the target values. This was the reason why it was decided to improve machine learning performance by automated features engineering implementation. On the basis of the original engine dataset, augmented engine health status dataset was created and new 35 predictors were generated.

In figure 9 results achieved of the machine learning algorithm engine health status prediction was presented. As it might be noticed, in this case scenario, the Predicted engine health status was definitely more accurate, and engine degradation was reflected with an error close to zero values.

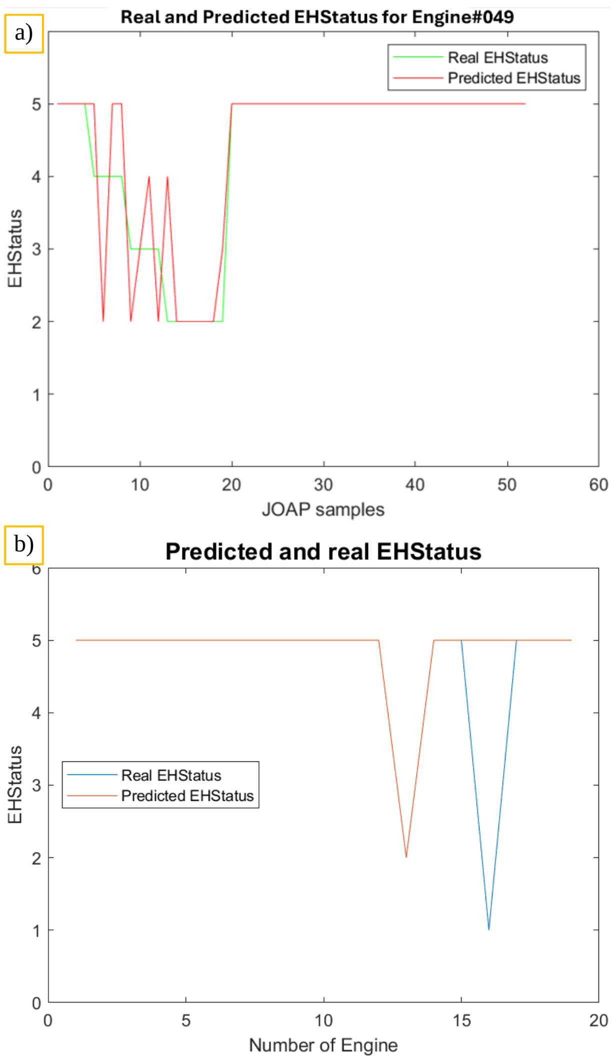

Fig. 8. Real and predicted Engine Health Status for Engine #049 (a) and plot of the predicted and real engine health status for all engines and original engine dataset (b).

In some cases, for instance for the MAPE, MSE and RMSE the improvement was greater of the order of magnitude. As it might be noticed from figure 9b for all the engines and augmented dataset the predicted values exactly reflect the real engine health status.

On the basis of the results achieved from the augmented machine learning classification, it turned out that two of the turbofan engines #049 and #052 should be set either on surveillance status on classified as unsafe for flight operations and removed from service. For both engines the alert for the propulsion maintenance management personnel should be triggered.

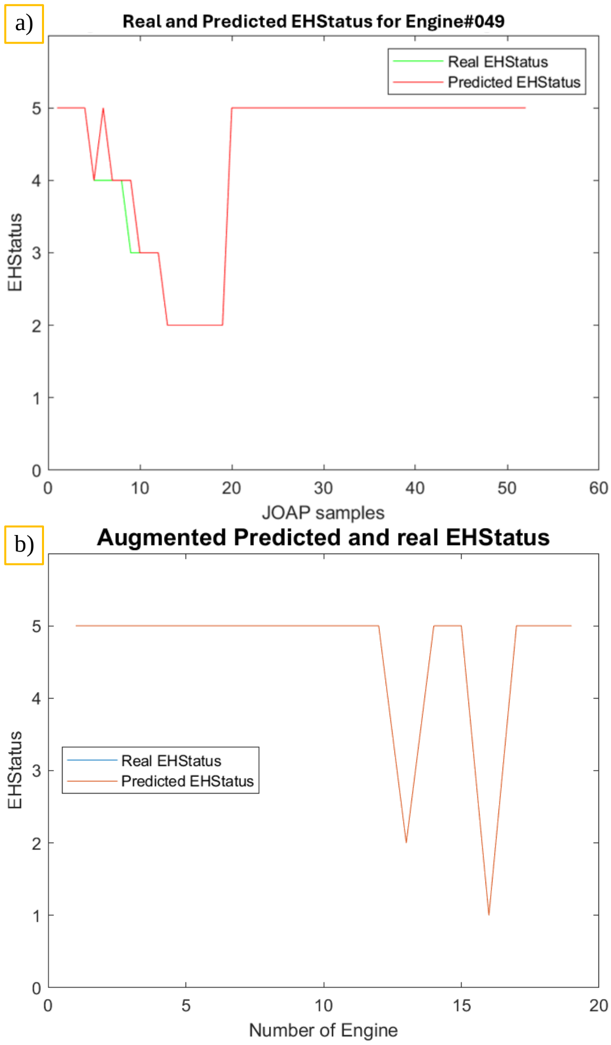

Fig. 9. Real and predicted Engine Health Status for Engine #049 (a) and plot of the predicted and real engine health status for all engines and augmented engine dataset with automated features engineering (b).

What should be the procedure for the engine with predicted Engine Health Status below OK Status. In the case when the concentration of any of the tracked elements exceeds the Abnormal Level then the Engine Health Status is classified as Engine Failure. If the trending concentration of the Fe and Ti elements (within the 10 last flight cycles) exceeds the abnormal level, engine must be set on surveillance for the next 10 flight cycles. If the Fe and Ti concentrations again exceed the abnormal (failure) level, then engine must be removed from the aircraft and flight operations and disassembled. While on surveillance, engine maintenance management should be focused on Ti concentration and remove engine from service immediately when the level of concentration reaches the Engine Failure Level. For the Fe, the level of concentration for the next two following flights must not exceed 5 ppm.

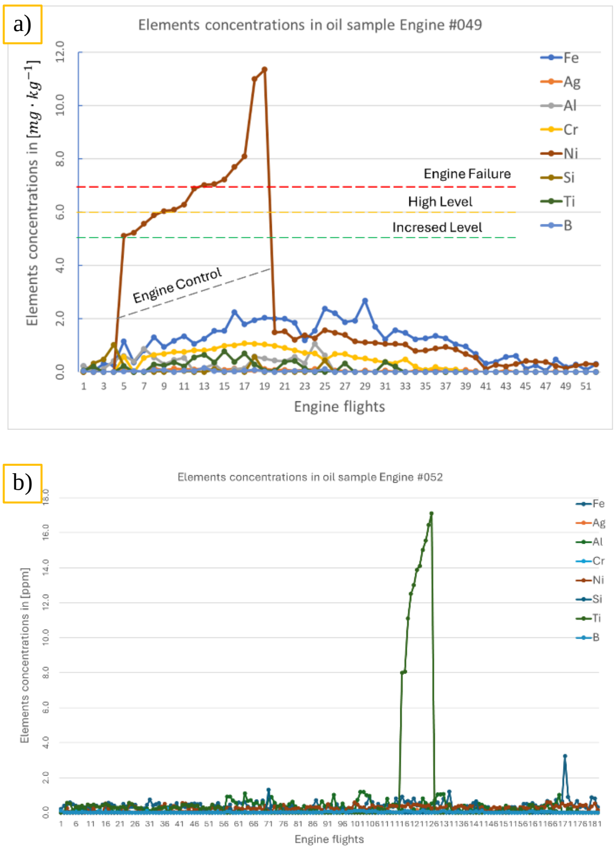

Fig. 10. Elements concentrations in oil samples for engine #049 (a) and #052 (b).

In figure 10 concentrations of the elements for the suspected engines #049 and #052 were presented. As it might be deduced from fig.10a the level of Ni concentration in oil sample collected from engine #049 started to increment from sample 4. If the increment trends exceed 2 ppm for the next 10 following engine flight hours, Engine Health Status must be changed to Engine Control. For the same engine oil sample, the Increased Level has been exceeded. The increment continued until for engine oil sample 9 reached High level and for sample 13 exceeded Engine Failure level. Engine #052 has been set on surveillance since engine oil sample number 113. The reason for this was the fact that the trending increment of Ti concentration exceeded 4 ppm. After two consecutive flight cycles the level of concentration reached Increased Level and continued to increase to reach High and Failure levels within the next 5 flight cycles.

7. SUMMARY AND CONCLUSIONS

Machine Learning algorithms might be a very strong and efficient tool in engine health status prediction based on the tribological data. What is worth mentioning is the fact that sometimes during engine flight operations there might be no sudden (rapid) increments in the amount of elements concentrations. What might be the real case scenario is the slow engine degradation. This situation could be detected with the analysis of the oil samples spectrometric results trending. It means that not only the level of the elements' concentration is important while predicting engine health status but also the trends of increments.

The main goal of the article was to present an idea of the artificial intelligence implementation for the tribology test results interpretation. The novel approach has been proposed and the whole methodology was discussed in a very precise way. On the basis of the collected engine oil sample test results, the augmented collection of the engine oil features was generated. This kind of features engineering allowed for the significant improvement of the machine learning predictions of the engine health status. Analyzing various machine learning performance metrics, it was possible to precisely compare predictions accuracy as well as the prediction errors. The predictions were also graphically presented, what allowed to visualize how the applied machine learning algorithm worked for all the engines.

The main goal of the article was achieved and confirmed for two of the engines. For both of them the proposed algorithm correctly confirmed the alerting change of the Engine Health Status. The machine learning algorithm not only indicated the alerting change of the engine health status, but also pointed which of the statuses were reached.

This allowed to, with reference to the designed Engine Health Status Model to focus on some of elements. On the basis of the analysis of the engine oil samples dataset it was detected that the degradation of the Engine Health Status was caused by the increment of the Ni element in engine #049 and Ti for engine #052.

On the basis of the case study and proposed machine learning methodology, it was decided to remove engine from service and check the condition of the suspected reasons for the elements concentrations increments. Analyzing construction of the case study turbofan engine as well as the type of the materials used to build engine components and parts the primary components to check should be as follows:

- engine #049 – No. 4 engine bearing compartment of the engine core module and high-pressure spool,
- engine #052 – No. 5 bearing compartment of the low-pressure turbine module.

Another extremely important factor of the research case study presented in the article was the fact that all the data was collected for the real turbofan engines. Engine oil sample test results were not generated or simulated but were collected for a few years of engine flight operations.

The question might be raised why building the engine health index number, which is used to define engine health status, only the presence of unusual concentrations of the elements in the fluid sample was taken under consideration. There are some publications where oil physical properties were considered.

For example, Jadhao [24] used machine learning methods to process high-dimensional simulation data generated in non-equilibrium molecular dynamics simulations, and ultimately obtained the correlation between rheological properties and molecular arrangement evolution in elastohydro dynamic lubrication, revealing the mechanism of viscosity decreasing with rates under low pressure of lubricants. Jones et al. [25] proposed one of the first applications of an ML algorithm in wear volume prediction for metal and alloys. They used load, speed, sliding distance, temperature, friction coefficient and kinematic viscosity as the input variables for machine learning algorithms. Peric et. al. [26] pointed out that if there is wear of the contact surfaces, there are wear particles present. Experimental results showed a change in the viscosity index of lubricants. Neural networks have also been applied by Afrand et.al [26] to lubricated contacts to model friction and lubricant viscosity to aid lubricant design.

Unfortunately, none of the physical properties is being tested on the object of the research being the aircraft turbofan engine. According to the engine manufacturer recommendations, engine oil physical properties have not been tested and tracked. This was the reason why oil physical properties were not taken under consideration while designing engine health status model. In addition to this, contrary to the mentioned publications, in the proposed methodology, we are not trying to predict engine oil life but the engine mechanical wear and degradation.

To summarize the results, it is possible to conclude that the machine learning algorithms might work well for the engine health status prediction based on the tribological data. Still, even for this number of features, the augmented engine oil sample dataset should be generated to achieved desired prediction accuracy and to minimize prediction errors.

As a future and following work the hybrid engine health status model could be developed which combines proposed machine learning algorithm and simulation model with tribological parameters like friction force, friction coefficient, contact temperature, noise, or vibration.

Presented methodology could be implemented into the aircraft (engine) maintenance management computer system, which could allow to automate engine health status analysis and improve engine maintenance management.

REFERENCES

[1] K. Feng, J. C. Ji, Y. Zhang, Q. Ni, Z. Liu, and M. Beer, "Digital twin-driven intelligent assessment of gear surface degradation," *Mechanical Systems and Signal Processing*, vol. 186, p. 109896, Nov. 2022, doi: 10.1016/j.ymssp.2022.109896.

[2] K. Feng, J. C. Ji, Q. Ni, and M. Beer, "A review of vibration-based gear wear monitoring and prediction techniques," *Mechanical Systems and Signal Processing*, vol. 182, p. 109605, Jul. 2022, doi: 10.1016/j.ymssp.2022.109605.

[3] M. Marian and S. Tremmel, “Current Trends and Applications of Machine Learning in Tribology—A review,” *Lubricants*, vol. 9, no. 9, p. 86, Sep. 2021, doi: 10.3390/lubricants9090086.

[4] A. Rosenkranz, M. Marian, F. J. Profito, N. Aragon, and R. Shah, “The use of Artificial Intelligence in Tribology—A Perspective,” *Lubricants*, vol. 9, no. 1, p. 2, Dec. 2020, doi: 10.3390/lubricants9010002.

[5] M. S. Hasan, A. Kordijazi, P. K. Rohatgi, and M. Nosonovsky, “Triboinformatics Approach for friction and wear prediction of Al-Graphite composites using machine learning methods,” *Journal of Tribology*, vol. 144, no. 1, Mar. 2021, doi: 10.1115/1.4050525.

[6] U. M. R. Paturi, S. T. Palakurthy, and N. S. Reddy, “The Role of Machine Learning in Tribology: A Systematic review,” *Archives of Computational Methods in Engineering*, vol. 30, no. 2, pp. 1345–1397, Nov. 2022, doi: 10.1007/s11831-022-09841-5.

[7] M. H. Rahman, S. Shahriar, and P. L. Menezes, “Recent progress of machine learning algorithms for the oil and lubricant industry,” *Lubricants*, vol. 11, no. 7, p. 289, Jul. 2023, doi: 10.3390/lubricants11070289.

[8] A. T. Sose, S. Y. Joshi, L. K. Kunche, F. Wang, and S. A. Deshmukh, “A review of recent advances and applications of machine learning in tribology,” *Physical Chemistry Chemical Physics*, vol. 25, no. 6, pp. 4408–4443, Jan. 2023, doi: 10.1039/d2cp03692d.

[9] A. Rosenkranz, M. Marian, F. J. Profito, N. Aragon, and R. Shah, “The use of Artificial Intelligence in Tribology—A Perspective,” *Lubricants*, vol. 9, no. 1, p. 2, Dec. 2020, doi: 10.3390/lubricants9010002.

[10] M. Marian and S. Tremmel, “Current Trends and Applications of Machine Learning in Tribology—A review,” *Lubricants*, vol. 9, no. 9, p. 86, Sep. 2021, doi: 10.3390/lubricants9090086.

[11] M. Marian and S. Tremmel, “Physics-Informed Machine Learning—An emerging trend in tribology,” *Lubricants*, vol. 11, no. 11, p. 463, Oct. 2023, doi: 10.3390/lubricants11110463.

[12] D. Chen, Y. Li, K. Liu, and Y. Li, “A physics-informed neural network approach to fatigue life prediction using small quantity of samples,” *International Journal of Fatigue*, vol. 166, p. 107270, Sep. 2022, doi: 10.1016/j.ijfatigue.2022.107270.

[13] Y. Xi, J. Deng, and Y. Li, “A solution for finite journal bearings by using physics-informed neural networks with both soft and hard constrains,” *Industrial Lubrication and Tribology*, vol. 75, no. 5, pp. 560–567, May 2023, doi: 10.1108/ilt-02-2023-0045.

[14] M. Rom, “Physics-informed neural networks for the Reynolds equation with cavitation modeling,” *Tribology International*, vol. 179, p. 108141, Jan. 2023, doi: 10.1016/j.triboint.2022.108141.

[15] Q. Ni, J. C. Ji, B. Halkon, K. Feng, and A. K. Nandi, “Physics-Informed Residual Network (PIResNet) for rolling element bearing fault diagnostics,” *Mechanical Systems and Signal Processing*, vol. 200, p. 110544, Jul. 2023, doi: 10.1016/j.ymssp.2023.110544.

[16] Y. Zhao, L. Guo, and P. P. L. Wong, “Application of physics-informed neural network in the analysis of hydrodynamic lubrication,” *Friction*, vol. 11, no. 7, pp. 1253–1264, Sep. 2022, doi: 10.1007/s40544-022-0658-x.

[17] Y. Cheng, Q. He, W. Huang, Y. Liu, Y. Li, and D. Li, “HL-nets: Physics-informed neural networks for hydrodynamic lubrication with cavitation,” *Tribology International*, vol. 188, p. 108871, Oct. 2023, doi: 10.1016/j.triboint.2023.108871.

[18] M. S. Hasan and M. Nosonovsky, “Triboinformatics: machine learning algorithms and data topology methods for tribology,” *Surface Innovations*, vol. 10, no. 4–5, pp. 229–242, Mar. 2022, doi: 10.1680/jsuin.22.00027.

[19] J. Walker *et al.*, “Application of tribological artificial neural networks in machine elements,” *Tribology Letters*, vol. 71, no. 1, Nov. 2022, doi: 10.1007/s11249-022-01673-5.

[20] N. Wang and C.-M. Tsai, “Assessment of artificial neural network for thermohydrodynamic lubrication analysis,” *Industrial Lubrication and Tribology*, vol. 72, no. 10, pp. 1233–1238, Jun. 2020, doi: 10.1108/ilt-03-2020-0109.

[21] M. S. Hasan, A. Kordijazi, P. K. Rohatgi, and M. Nosonovsky, “Triboinformatics Approach for friction and wear prediction of Al-Graphite composites using machine learning methods,” *Journal of Tribology*, vol. 144, no. 1, Mar. 2021, doi: 10.1115/1.4050525.

[22] R. Liu, B. Yang, E. Zio, and X. Chen, “Artificial intelligence for fault diagnosis of rotating machinery: A review,” *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, Feb. 2018, doi: 10.1016/j.ymssp.2018.02.016.

[23] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, “Deep learning and its applications to machine health monitoring,” *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, Jun. 2018, doi: 10.1016/j.ymssp.2018.05.050.

[24] V. Jadhao and M. O. Robbins, “Rheological properties of liquids under conditions of elastohydrodynamic lubrication,” *Tribology Letters*, vol. 67, no. 3, May 2019, doi: 10.1007/s11249-019-1178-3.

[25] S. P. Jones, R. Jansen, and R. L. Fusaro, “Preliminary investigation of neural network techniques to predict tribological properties,” *Tribology Transactions*, vol. 40, no. 2, pp. 312–320, Jan. 1997, doi: 10.1080/10402009708983660.

[26] S. Peric, B. Nedic, and G. Aleksandar, “Applicative monitoring of vehicles engine oil,” *Tribology in Industry*, vol. 36, no. 3, pp. 308–315, Sep. 2014.

[27] M. Afrand *et al.*, “Prediction of dynamic viscosity of a hybrid nano-lubricant by an optimal artificial neural network,” *International Communications in Heat and Mass Transfer*, vol. 76, pp. 209–214, May 2016, doi: 10.1016/j.icheatmasstransfer.2016.05.023.