Acoustic emission and machine learning based classification of wear generated using a pin-on-disc tribometer equipped with a digital holographic microscope

Pushkar Deshpande *, Vigneashwara Pandiyan, Bastian Meylan, Kilian Wasmer

Laboratory for Advanced Materials Processing (LAMP), Empa - Swiss Federal Laboratories for Materials Science & Technology, Unterstrass 39, CH-3602, Thun, Switzerland

**ARTICLE INFO**

**Keywords:**
- Machine learning
- Acoustic emission
- Wear prediction
- In-situ monitoring
- Digital holographic microscope

**ABSTRACT**

The efficiency of processes involving frictional contacts between surfaces is often characterized by wear rates or friction coefficients. However, the classification and forecasting of wear rates in friction related processes is a real industrial challenge that is unsolved today. Hence, an on-line monitoring system able to classify wear rate can be crucial for many industries as it could help in preventing catastrophic failures. Applications include lifetime assessment of various industrial components where a range of wear failures occur such as scuffing (a typical sudden failure mechanism). These tribological processes can now be sensorized, and the corresponding sensor signatures can be modelled and monitored using state-of-the-art Machine learning (ML) algorithms. In this study, we use an Acoustic Emission (AE) sensor and ML frameworks to classify different wear categories simulated with a customized pin-on-disc tribometer. A real-time investigation of the wear track is necessary to find out the origins of the wear scar visible at the surface. To achieve this objective, the experiments were conducted on a pin-on-disc tribometer equipped with a Digital Holographic Microscope (DHM). Experiments were carried out using alumina and steel balls against steel discs at room temperature. Real-time DHM images of the wear track surface were recorded for each lap at the same position. An acoustic emission sensor recorded the AE signals during the complete duration of experiments. The AE signatures, in combination with the real-time DHM images, were correlated as input and ground truth labels for the ML algorithm. Several ML frameworks were compared; they are support vector machine, logistic regression, XGBoost, random forest, neural networks, k-Nearest Neighbor, quadratic discriminant analysis and Naive Bayes. The classifier was trained to differentiate the acoustic emission features of the different wear rates. Most ML algorithms had an average classification accuracy above 80%, and the highest was obtained with support vector machine (84.7%). The classification accuracy can be improved by combining two neighboring categories with limited differences in terms of wear rate. Hence, the proposed method has a significant industrial potential for in-situ and real-time quality monitoring of wear processes since it requires minimum modifications of commercially available industrial machines.

1. Introduction

Direct contact between solids surfaces in industrial components causes failures such as abrasive and/or adhesive wear depending on the type of contact (dry or lubricated) [1,2]. Abrasive wear occurs predominantly when a hard surface or a particle is rubbed against a soft surface [3]. This is commonly observed in various industrial processes such as mining, drilling, grinding, digging, starved lubrication conditions and wherever third body particles are involved in lubricated solid contacts [4]. Under these circumstances, wear debris is generated in these contacts, and this acts as a third body abrasive; which affects the wear rates considerably depending on its size and properties. The choice of materials for industrial applications is often based on wear resistance, and so there is a need to quantify the wear characteristics. The wear in tribological contacts are estimated by measuring the weight loss of the materials in contact, wear rates [5], wear volumes and by observing changes in the wear scar width and depth [5–9].

Post-mortem quantification of wear rates is trivial as they differ across all frictional processes. The conventional post-process wear quantification techniques are profilometry, confocal microscopy and
interferometry. Unfortunately, these techniques do not give any insight into the dynamics of the wear process [10]. To have access to such information and to understand the different transients between the initial and the final surface states, it is necessary to measure the wear in real-time. In other words, only an in-situ and real-time monitoring system will enable to detect abnormal and/or various wear behavior during the process. This also gives the opportunity to link the data acquired to the mechanisms. Wear monitoring is essential for a variety of industry contacts to identify the onset of severe wear to stop the process or to repair/replace the components [11]. Besides, the prediction of remaining useful lifetime of components could also be possible [12]. Real-time wear detection and monitoring will also strengthen the various hypotheses proposed to wear mechanisms for different material pairs in the academic community.

Real-time wear monitoring and measurement include direct and indirect methods [10]. In the case of direct methods, wear is measured by electrical, optical or radioactive resistance sensors. Examples of optical methods include measurement of surface topography of the wear scar using microscopic techniques like Digital Holographic Microscope (DHM) [11] or Scanning Electron Microscope (SEM) [13]. Indirect methods of wear monitoring/measurement include wear evaluation based on parameters measured during Acoustic Emission (AE), Ultra (DHM) [11] or Scanning Electron Microscope (SEM) [13]. Indirect methods include measurement of surface topography of the wear scar using microscopic techniques like Digital Holographic Microscope (DHM) [11] [14–16]. AE techniques are currently the most widely used for monitoring and analysis of wear for a variety of frictional contact processes [17]. AE is based on the generation of acoustic waves, signatures from the solid materials in contact whenever they are stressed and relaxed [14]. Furthermore, AE signals can be recorded with acoustic emission sensors and then analyzed to generate useful information about the physics of the various wear processes [18–20].

AE has been widely used to monitor wear for various laboratory scale as well as industrial-scale frictional failures such as scuffing, fretting, rolling contact, fatigue, etc. [18–27]. Jun Sun et al. studied wear monitoring of bearing steel on a pin-on-disc tribometer under dry sliding conditions with a combination of acoustic and electrostatic emission [16]. They were able to monitor various phases of delamination wear. Bones et al. showed that the time-dependent nature of the AE signal could detect the presence of anti-wear additives and predominant wear processes [28]. Sarychev et al. exploited AE signals in frictional processes to indicate the frictional state, quantity of solid and liquid layers in contacting surfaces in real-time as well as other information such as the intensity changes of wear and friction [29]. Literature also shows that AE signals could be generated from non-metallic materials, in particular, polymers, e.g. polyether ether ketone (PEEK) and plasma-sprayed coatings to monitor wear [15,29].

Acoustic emission signal analysis is a well-known method for identifying wear based on the frequency signatures from different materials in order to classify and predict wear accurately [14,30,31]. Obviously, owing to the high temporal resolution of the AE, vast amounts of data are generated during AE experiments. However, few studies have been performed on the classification of different lubrication regimes and prediction of wear rates using AE signals and Machine Learning (ML) algorithms. Different methods are considered to analyze the acquired AE signals. They include statistical methods such as regression analysis and recent advanced machine learning methods; e.g. Support Vector Machine (SVM), Artificial Neural Network (ANN) [8,12,32,33] etc. The significant research works carried out in various frictional processes to classify and predict wear by machine learning methods are summarized in Table 1.

Table 1

<table>
<thead>
<tr>
<th>Simulated friction process</th>
<th>Sensor</th>
<th>Machine learning algorithm</th>
<th>Research group</th>
<th>Publication year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wear of journal bearings</td>
<td>AE</td>
<td>SVM</td>
<td>Mohktari et al.</td>
<td>2020</td>
</tr>
<tr>
<td>Scuffing prediction, failure in lubricated surfaces</td>
<td>AE</td>
<td>Random forest</td>
<td>Wasmer et al.</td>
<td>2017</td>
</tr>
<tr>
<td>Scuffing identification and prediction</td>
<td>Force</td>
<td>Recurrent neural networks</td>
<td>Tyler et al. [36]</td>
<td>2017</td>
</tr>
<tr>
<td>Bearing remaining useful life</td>
<td>AE</td>
<td>SVM</td>
<td>Elforjani et al.</td>
<td>2015</td>
</tr>
<tr>
<td>Classification of frictional wear, Wear detection</td>
<td>AE</td>
<td>SVM, ANN</td>
<td>Olorunlambe et al.</td>
<td>2019</td>
</tr>
<tr>
<td>Biological joints – fretting wear</td>
<td>AE</td>
<td>ANN, GA</td>
<td>Sadegh et al.</td>
<td>2015</td>
</tr>
<tr>
<td>Abrasive wear rates prediction</td>
<td>Force</td>
<td>LR &amp; Random forest</td>
<td>Kalentiev et al.</td>
<td>2018</td>
</tr>
<tr>
<td>Predicting &amp; analyzing abrasive wear rates</td>
<td>Force</td>
<td>ANN</td>
<td>Thankachan et al.</td>
<td>2020</td>
</tr>
<tr>
<td>Prediction of wear loss of Fe Alloys</td>
<td>Force</td>
<td>SVM, LR and GPR</td>
<td>Altay et al. [43]</td>
<td>2020</td>
</tr>
<tr>
<td>Prediction of wear loss in Mo coatings</td>
<td>Force</td>
<td>ANN</td>
<td>Cetinel et al.</td>
<td>2006</td>
</tr>
<tr>
<td>Prognosis of bearings</td>
<td>AE</td>
<td>SVM, LR and GPR</td>
<td>Elforjani et al.</td>
<td>2018</td>
</tr>
</tbody>
</table>

The reconstructed DHM images are used as the ground truth to divide the friction curve into five different wear categories. Acoustic features calculated from the domains such as time, frequency and time-frequency are labelled to the corresponding five wear categories. The feature corresponding to the five wear categories is introduced to various supervised machine learning classification algorithms. The robustness of the methodology is verified by comparing various machine learning classification algorithms.

The paper is organized into four sections. Section 1 presents an overview of wear mechanisms and monitoring techniques based on ML. Section 2 discusses the tribological conditions, experimental setup, wear characteristics and ML framework. Section 3 reports the results of the feature analysis, t-distributed stochastic neighbor embedding (t-SNE) visualization and classification accuracy using SVM. Finally, the conclusions of this research work are reviewed and discussed in Section 4.

2. Experimental methods

In this section, experimental methods are presented including materials, rigs, test conditions and characterization techniques as well as the machine learning framework.

2.1. Pin-on-disc tribometer with DHM for real-time surface characterisation of the wear scar

The tribometer used in this work is a pin-on-disc customized high vacuum high-temperature tribometer developed by Anton Paar Tri Tec SA, as shown in Fig. 1. TriboX software controls all tribometer axis movements [11].
The tribometer is additionally equipped with an in-situ DHM developed by Lyncee Tec SA. The DHM allows pseudo real-time imaging of the wear track. A dual-wavelength (2 laser wavelengths of 415 ± 485 nm) customized.

DHM was selected for this purpose. Holograms are acquired with a 1.4 Megapixel CMOS camera, which can be synchronized with the rotation of the disc holder via an external trigger. This ensures successive image acquisitions at the same position for each lap, enabling long term monitoring of the wear track using the DHM Koala software [11].

2.2. Materials and test conditions

The materials selected for the tribotests were mirror-polished steel discs (100Cr6) and alumina balls (6 mm diameter) pairs. The average surface roughness (Ra) for the steel discs prior to the experiment was 25 ± 15 nm and 20 ± 10 nm for the alumina balls. The ball (pin) holder was the stationary part of the tribometer, and it was equipped with a load cell to apply the contact force. The tribotest conditions were selected to ensure that abrasive wear occurred on the surface of the steel disc. However, adhesive wear could also be present but to only a small extent. The experiment was carried out at a constant load of 1 N at ambient temperatures. Two tests were conducted with the same conditions to ensure the repeatability. As the steel disc and alumina balls were standard materials, elastic moduli provided by the manufacturer was used to calculate the Hertzian contact pressure. The elastic moduli of steel was 210 GPa, and that of alumina was 350 GPa. Therefore, according to the Hertzian theory of contact mechanics [44], the maximum contact pressure was 770 MPa. The rotational speed of the disc was set at 10 cm/s, and the total sliding distance was 100 m. The tribotest conditions are summarized in Table 2.

2.3. Abrasive wear monitoring – acoustic emission data acquisition

Acoustic emission signals were recorded with a standard commercial Vallen acoustic emission acquisition system. Signals were collected in a continuous mode for the total duration of the experiment. The acquisition was made with a GF Nano 30 miniature AE sensor from Physical Acoustics (PAC). It was fixed with a polymer glue on to the stationary ball holder to ensure that it remained stable for the complete duration of the experiment. The polymer adhesive allows good transmission of AE signals and was easily detachable after the test. The distance between the AE sensor and the sliding surface was around 4 cm. The sensor has a frequency response over the range of 125–750 kHz, and its peak sensitivity was 72 dB. This AE sensor was chosen, as mechanical wear mechanisms are known to occur in low-frequency ranges < 1000 kHz [19]. The sampling rate during acoustic emission data acquisition was 2 MHz. The amplitude of the AE signals generated during the friction test is low, and therefore, an amplifier with an integrated bandpass filter (2/4 preamplifier) was used. Two amplifier stages of 20 and 40 dB could be set. However, in this work, a 40 dB gain was selected as the friction signals have low amplitudes. This system consists of an in-built bandpass filter (bandwidth of 100–1000 kHz) which attenuates the extremely low and high frequencies generated during the experiment.

The raw sensor data was split into fixed-width sliding windows of 2500 µs. At a sampling rate of 2 MHz, each window of 2500 µs contains 5000 points. A vector of several features corresponding to the time, frequency and time-frequency domains were extracted from each raw sliding window. Then, they were labelled into five categories based on the DHM ground truth wear images. The wear categories were selected according to the percentage of wear observed above the theoretical Hertzian wear scar width and from the reconstructed DHM images (depth profiles). The feature set was optimized by eliminating the non-informative and redundant features by recursive feature elimination based on logistic regression. Next, the new feature subset was introduced into the supervised machine learning classification algorithm; mainly an SVM framework. The model was trained with 75% of subset feature dataset, and the remaining 25% was used for testing. This approach simulated real-life conditions where the trained system has to operate with new input data. The robustness of the trained SVM model was verified by comparing the prediction of the model on the experimental set with the ground truth labels. This framework to classify wear is described in Fig. 2.

Several ML frameworks have been selected to compare the SVM results and test the robustness of the approach. They were logistic regression, XGBoost, Random Forest, neural networks, k-Nearest Neighbor, quadratic discriminant analysis and Naive Bayes.

3. Results and discussions

3.1. Friction results

A representative friction curve for alumina ball on steel disc contact tribotest in the dry condition is shown in Fig. 3. The friction curve is divided into three stages: running-in, steady-state and severe abrasive wear failure. The first stage shown in red in Fig. 3 from 0 to 400 s (0–200 laps) is defined as running-in and known from previous research works [45]. During this stage, the direct contact between the polished hard
alumina ball and the softer steel disc makes the surface of the latter roughened. This marks the beginning of abrasive wear [3,46]. After the initial running-in, the friction coefficient fluctuates a little and then remains constant in the range of 0.5–0.6 between 400 and 900 s. This stage is labelled as the steady-state stage (201–450 laps) shown in green in Fig. 3 where roughening of the polished steel surface continues and the abrasive wear goes on increasing progressively. Wear debris is generated during this stage, but it does not affect the wear or friction coefficient considerably. After the steady-state stage, severe third body abrasive wear occurs due to the debris in the contact from 900 to 2000 s. This stage is labelled as severe abrasive wear failure (451–1000 laps). In this stage, the friction coefficient fluctuates regularly between 0.2 and 0.9 and severe third body abrasive wear occurs due to the wear debris in the contact. The fluctuations in the friction coefficient are due to the steel debris moving in and out of the contact making the surface soft or hard. Depending on the generated contact pressures, friction coefficient changes (lower pressures – lower friction coefficient and higher pressures – higher friction coefficients). The five different wear categories are also shown in Fig. 3, and these are discussed later in Section 3.2.

3.2. Wear categorization and real-time DHM images and profiles of the wear scar

It is obvious from Fig. 3 that the wear categories do not correspond directly to the running-in, steady-state and severe wear stages. The reason is the classification task of the later stage has already been carried out [21,31]. Consequently, in this work, five wear categories were made based on percentage wear above theoretical Hertzian diameter (58 μm). The category 1 to category 5 are defined as < 100% wear, < 150% wear, < 200% wear, < 300% wear and > 300% wear, respectively. The defined categories were later labelled and confirmed from the processed DHM images and the wear scar depth profiles as the ground truth.

As discussed in the experimental methods in Section 2.1, real-time DHM images (holograms, phase and intensity images) were acquired for each lap at a specific location during the tribotest. The raw DHM holograms at specific laps in different wear categories are shown in Fig. 4 (a). The DHM phase images were processed using the Koala software from Lynceetec SA. From the recorded phase images, an image of the cross-section of the wear scar was reconstructed digitally. It was observed that the reconstruction of the phase images obtained for the few initial laps was relatively easier as the abrasive wear was just initiated and was minimal.

The images in Fig. 4 (a) are processed and reconstructed, and the results are shown in Fig. 4 (b). From these images, it can be observed that the raw images for laps 253, 447 and 880 are incomplete, and this is due to extreme wear and particles stuck in the scratches of the wear scar. However, from the DHM processed phase images, we could measure the depth profiles accurately. In Fig. 4 (b) lap 880, the wear scar width is extremely large, and it almost crosses the maximum possible lateral width visible in an image which is around 200–220 μm.

A typical wear scar depth profile per wear category is shown to highlight the differences in the wear scar width and depth in Fig. 5. It is evident that the width and depth of the wear scars increase considerably as the wear categories change. The depth profile for lap 12 (wear category 1) shown in Fig. 5 (red line scan) is minimal whereas the wear scar width is almost equal to the theoretically calculated Hertzian diameter of 58 μm [47]. This wear scar width is found to be well below 100 μm. In this wear category, the scratches and material removal occurring in this category is attributed to the initial running-in period. This is confirmed by the reconstructed 3-dimensional (3D) image shown in Fig. 6 (lap 12). The 3D DHM images of the wear scars in Fig. 6 are with a scale of 209 by 209 μm. The wear scar depth profile for lap 107 in the wear category 2 in Fig. 5 (brown line scan) clearly shows an increase of the scratches depth as well as material removal. The scratch generation is manifested from the 3D image in Fig. 6 (lap 107). As the sliding continues, particles are removed from the surface of the steel disc and the already scratched solid material surface is influenced by a third body abrasive – steel particles. This was confirmed by a visual inspection of the alumina ball, showing no sign of wear. The presence of the third body increases the abrasive wear considerably. This is obvious from the friction coefficient fluctuations observed in Fig. 3 (400–2000 s) as the particles move in and out of the contact.
In the wear category 3, the wear scar depth profile for lap 253 (green line scan) in Fig. 5 shows a different profile surface. On one half, there are deeper and wider scratches whereas, on the other half, the surface is quite flat. This is also corroborated by the 3D reconstructed image, and evidence of this is in Fig. 6 (lap 253). In this case, the wear scar width is increased to more than 150% above the theoretical Hertzian width. Further profiles in the wear category 4 and category 5, laps 447 and 880, respectively, are significantly different but both show an important augmentation of the wear depths to more than the bearable wear limits and the scratches have a depth ranging from around 2 to 5 μm. In addition, the wear scar widths are between 150 and 200 μm (more than 200 or 300% wear as compared to the theoretical wear scar depth). The 3D images shown in Fig. 6 confirm the high wear scar depths and widths for the laps 447 and 880 in the wear category 4 and category 5, respectively. As observed from these profiles and DHM images, it is clear that contact surface area changes (due to the increase in wear scar width and depth) during the test as the wear increases. However, as already mentioned, there is no observable sign of wear on the alumina ball after the test. Therefore, the changes in wear are solely related to the wear of the abrasive wear of the steel discs (100Cr6).

The five wear categories defined previously are shown in Table 3. These categories are labelled and confirmed on the basis of the DHM images, wear scar profiles and 3D reconstructed images in Fig. 4, Fig. 5 and Fig. 6. The wear categories are also marked on the friction curve in Fig. 3. It is clear that the divided wear categories are not exactly sensitive to the fluctuations in friction coefficients, but the real time DHM images which are used as ground truth play an important role in confirming the defined categories. Also, this suggests that analyzing the wear only based on the changes in the friction coefficient are not enough. Consequently, the use of acoustic wave signatures from the sample could provide with more details of the physics of the frictional process. Therefore, in the next subsection, AE signals are correlated with wear categories and discussed. Classification of the wear categories is performed using ML algorithms from the processed AE signals.

3.3. Acoustic emission signal analysis

The AE signal recorded is resolved into statistical features corresponding to three domains, namely time, frequency and time-frequency (wavelet). The filtering of the AE signals is a necessary step before the feature extraction. It is performed via a bandpass filter, which is already prebuilt in the Vallen data acquisition system. The signal to noise ratio for the AE signals was found to be around 25 dB. The feature extraction and their analysis are discussed in the next sections.

Feature extraction and analysis

Fig. 7 presents examples of raw AE signals for a window of 2500 μs or 5000 points (at 2 MHz) corresponding to the five defined wear categories. A visualization of the raw signal indicates that the AE signal fluctuates significantly with time. But, by comparing the different categories, no visual features typical for each category are distinguishable, making the classification task suitable for ML frameworks. A visualization of the raw signal indicates that the amplitude of the AE signals increases with the wear till category 4 and subsides after category 5. It is also observed that once the surface is re-smoothened around category 5; there is a considerable decrease in the AE amplitude as seen on the y-axis of the individual plots.
Time-domain analysis

The time-domain analysis gives the statistical features that are a function of the variation of the amplitude over time. Fig. 8 and Fig. 9 show the RMS distribution and the skewness features, respectively, of the AE signals for the five categories (2500 μs window size has 5000 data points). The RMS distributions between the five wear categories are distinct, as evident from Fig. 8. The skewness distribution, depicted in the form of violin plots, as shown in Fig. 9, reveals that the wear categories are skewed both positively as well as negatively. The distinct shape of the individual violin plot corresponding to the five wear categories is encouraging to use them as an input feature in the ML algorithms. The statistical distributions of the features based on time-domain such as mean, kurtosis, median, crest-power, etc. were also analyzed for the five selected wear categories, and it was found that they also had similar discrete trend between the categories.

Frequency domain analysis

The periodical patterns that repeat itself with time in a signal can be identified when it is moved from the time domain to the frequency domain. This transformation to the frequency domain is achieved through the spectrum analysis of the AE signals obtained for the five different wear categories during the tribotest. This would be enhanced by using advanced signal processing techniques like wavelet transform, which can provide a time-frequency representation of the signal. The frequency domain analysis was performed using the Fast Fourier Transform (FFT) technique. The obtained FFT results were then visualized in a 3D bar graph for each wear category, as shown in Fig. 10. The bar graphs displayed the frequency content of the AE signals for each wear category, with the X-axis representing frequency, the Y-axis representing amplitude, and the Z-axis representing the number of occurrences. The bar graphs clearly showed that the AE signals for different wear categories had distinct frequency distributions, indicating a change in the wear mode of the contact surfaces. This frequency-domain analysis can be further extended using advanced signal processing techniques like frequency-domain analysis, which can provide a detailed insight into the wear mechanisms.
domain. Decomposing a signal in a Fast Fourier Transforms (FFT) analysis reveals the details of these periodical components. For the AE signals acquired from the tribotest, distinct peaks were found between 100 and 300 kHz and evidence of this is seen in Fig. 10 across the five wear categories. The Power Spectral Density (PSD) distribution in the FFT plots was calculated using the Welch method [48] for a window size of 2500 μs similar to the time-domain analysis.

To understand the correlation between the energy of the signal on the five wear categories, the frequency range was resolved equally into five energy bands between 0 and 300 kHz. The maximum frequency was fixed at 300 kHz, which corresponds to maximum distinctive peaks found from an FFT analysis. Hence, the five energy bands were 0–60 kHz, 60–120 kHz, 120–180 kHz, 180–240 kHz and 240–300 kHz. The results, in terms of cumulative energy values calculated for the five energy bands corresponding to each category, is presented in Fig. 11. The energies were calculated by the periodogram method [49]. From Fig. 11, we observed that the energies contents were concentrated within the frequency range of 120–240 kHz during the contacts of steel with the ceramic ball. In the case of wear category 5, where the wear scar width and depth are significantly high, and the wear is extreme, it is evident that the energy content in the energy band between 120 and 180 kHz was significantly higher than the other categories. The frequency plots and energy distribution of the AE signals between the different bands suggest that the categories can be statistically separated. Besides, the discrete distribution between the categories gives the motivation to use them as input for classification algorithms for real-time monitoring.

**Time-frequency domain analysis**

The resolution of the AE signals in the frequency domain along with its time localization can be carried out via Wavelet transformation (WT) [50]. Continuous Wavelet Transformation (CWT) was calculated on the filtered AE signals using Morlet [51] as mother wavelet with 500 scales. The 3D representation of the wavelet transforms are shown in Fig. 12. The wavelet coefficient values were predominantly in the range of 100–300 kHz for the AE signals irrespective of the category, which are synonymous to the analysis performed in the sub-section frequency domain analysis. The distribution of the wavelet energy coefficients for the wear categories suggests that the time-resolved frequency features computed using WT can be used to distinguish the five wear categories.

**t-SNE visualization**

Establishing correlations between numerous features in the higher dimension is an arduous task. However, representing the features in

---

**Fig. 8.** Distribution of AE RMS features for the five wear categories.

**Fig. 9.** Skewness distribution plots of AE signals corresponding to all the five wear categories.

**Fig. 10.** FFT plots for all the five wear categories.
lower dimensions helps to interpret the feature space. The dimensional reduction technique, such as t-distributed stochastic neighbor embedding (t-SNE), enables visualizing high-dimensional datasets in the lower dimension. The t-SNE is a nonlinear dimensionality reduction and visualization technique that maps the probability distribution of various neighboring points to a low dimensional space that follows a probability distribution as best as possible. As compared to other nonlinear dimensionality techniques, t-SNE is capable of retaining both the local and global structure of the data at the same time. The AE features from the time, frequency, and time-frequency domains listed in Table 4 were the inputs to perform the t-SNE computation. When performing the t-SNE computation, a perplexity value has to be selected. The perplexity is a hyper-parameter responsible for preserving the local and global structure of the data that determines the number of neighbors to be considered for embedding. In this study, a perplexity value of 10 was selected on an exhaustive search by visualizing the clusters in feature space. This value allows for understanding the feature space distribution of the AE signals recorded during the experiments, and the results are presented in Fig. 13. A movie showing different perspectives can be watched in the attached files with the submission. The t-SNE visualization results can be summarized as follows: the presence of clusters corresponding to the five wear categories in the feature space demonstrates that the features extracted from the three domains (time, frequency, time-frequency domains) can be potentially combined with nonlinear ML algorithms for precise classification.

Table 4

<table>
<thead>
<tr>
<th>Domain</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Mean, RMS, Kurtosis, Skewness, Crest factor, Standard deviation, Minimum, Maximum, Median etc.</td>
</tr>
<tr>
<td>Frequency</td>
<td>Position of peaks with high intensity, Energy distribution in respective energy bands</td>
</tr>
<tr>
<td>Time-frequency (Wavelet)</td>
<td>Enthalpy, RMS, Kurtosis, Skewness, Standard deviation in respective decomposition levels etc.</td>
</tr>
</tbody>
</table>
3.4. SVM classification

For classification tasks, building a classifier for the five wear categories consists of three main steps: feature reduction, training of the classifier and testing robustness of the algorithm. A set of 304 features as listed in Table 4 were extracted for the sliding window length of 2500 µs (5000 points) prior to building the classifier. First, a recursive feature selection based on logistic regression was carried out to pick a subset of 50 features from the original 304 features for training the classifier. Logistic regression-based feature selection technique tries to come up with a good subset of features (50 features in our case) iteratively on all possible combinations but without conceding on accuracy. The features that are redundant, non-informative and insignificant are eliminated at each iteration. The second step involves the training of the ML classifier. From the subset containing 50 features, 75% of the data is selected stochastically for the training and the other 25% for testing the developed models. Prior knowledge of the features representation for the five wear categories in t-SNE visualization gave hints to select the SVM as ML technique. The reason is that SVM is ideal for solving nonlinear classification problems via kernel trick with ease. Using kernel trick, linear separation can be achieved from nonlinear SVM by means of a mapping function [32,51,52]. SVM takes advantage of prior knowledge of the wear categories and constructs a hyperplane as a decision surface so that the periphery of the separation between the wear categories are separated. The Python open-source libraries SciPy [53] and scikit-learn [54] were used for achieving the necessary data transformation and training of the SVM classifier. The choice of hyper-parameters such as the regularization parameter (C) and the kernel parameter (γ) for the SVM classifiers was also chosen based on an extensive grid search. Finally, the robustness of the trained SVM classifier was verified with a test dataset.

The classification results from the SVM are shown in Table 6. In this table, the wear categories (categories 1–5) (in rows) versus the ground truth (in columns) are given. The classification accuracies in the table are defined as the number of true positives divided by the total number of tests for each category. These values are given in the diagonal cells of the table (highlighted grey cells). The classification errors are computed as the number of the true negatives divided by the total number of the tests for each category. These corresponding values are filled in non-diagonal row cells. Based on these results, the classification confidence for the wear categories varies in the range of 58–100%. These results clearly show the potential of the proposed approach, in particular when taking into account a non-optimized setup. Hence, we can conclude that the AE signal processing with SVM can be a versatile approach for in-situ and real-time wear monitoring method.

The analysis of the classification errors structure can be evaluated from the non-diagonal rows in Table 6. For example, the AE test signals from the wear category 1 were classified with an accuracy rate of 95%.

### Table 5

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kernel function</td>
<td>Radial basis function</td>
</tr>
<tr>
<td>Kernel scale</td>
<td>Automatic</td>
</tr>
<tr>
<td>Features</td>
<td>50 Subset features</td>
</tr>
<tr>
<td>Multiclass method</td>
<td>One-vs-One</td>
</tr>
<tr>
<td>Standardize data</td>
<td>True</td>
</tr>
<tr>
<td>Regularization parameter C</td>
<td>0.1,1,10,100</td>
</tr>
<tr>
<td>Kernel parameter γ</td>
<td>0.1,0.01,0.0001</td>
</tr>
</tbody>
</table>

The classification error is more or less equal between the wear category 2 and category 4, 3 and 2%, respectively. The average accuracy of the SVM is high as 84.8% with most errors occurring due to misclassification between category 3 and category 4, as evident in Table 6. This overlap is certainly due to the similarities in wear scar widths and depths and so in the AE signals. Based on this statement, the classification accuracy could be significantly increased by combining the wear category 3 and category 4. Actually, by doing so, the classification of the new category is as high as 96%. The fraction of the samples misclassified of the developed SVM model were minimum as depicted by the confusion matrix in Table 6.

### Table 6

<table>
<thead>
<tr>
<th>Predicted label (%)</th>
<th>Category 1</th>
<th>Category 2</th>
<th>Category 3</th>
<th>Category 4</th>
<th>Category 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground truth (true label)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category 1</td>
<td>95.0</td>
<td>3.0</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Category 2</td>
<td>3.9</td>
<td>96.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Category 3</td>
<td>3.0</td>
<td>0.0</td>
<td>58.0</td>
<td>39.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Category 4</td>
<td>2.5</td>
<td>1.5</td>
<td>21.0</td>
<td>75.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Category 5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>100.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

3.5. Comparison of the SVM with other ML algorithms

In the previous section, we presented the classification results obtained with the SVM. In order to verify the robustness of the approach as well as the results, a comparison with the main state-of-the-art ML algorithm is performed. The choice of the ML algorithms is given in Table 7, with their average classification accuracy. All confusion matrices for the ML algorithms are given in Table A1 in the Appendix. Scrutinizing the confusion matrices in Table A1, it is important to note that the observations made for the SVM are still valid for the other ML algorithms. This is particularly true for the misclassification between the wear category 3 and category 4. This supports the observation that both wear categories are extremely close in terms of wear scar widths and depths and so in the AE signals and not due to the classification method. Thus, only the results of the average classification accuracy are
Table 7
Prediction accuracies for the different machine learning classification algorithms.

<table>
<thead>
<tr>
<th>ML algorithms</th>
<th>Average classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support Vector Machines (SVM)</td>
<td>84.7%</td>
</tr>
<tr>
<td>Neural Networks (NN)</td>
<td>82.1%</td>
</tr>
<tr>
<td>Logistic regression (LR)</td>
<td>83.9%</td>
</tr>
<tr>
<td>Random Forest (RF)</td>
<td>83.1%</td>
</tr>
<tr>
<td>Naive Bayes (NB)</td>
<td>73.8%</td>
</tr>
<tr>
<td>k-Nearest Neighbor (k-NN)</td>
<td>81.0%</td>
</tr>
<tr>
<td>Quadratic Discriminant Analysis (QDA)</td>
<td>75.4%</td>
</tr>
<tr>
<td>XGBoost (XGB)</td>
<td>83.2%</td>
</tr>
</tbody>
</table>

compared. Based on Table 7, it is seen that SVM has the highest average classification accuracy (84.7%) followed closely by logistic regression (83.9%), XGBoost (83.2%), Random Forest (83.1%), neural networks (82.1%) and k-Nearest Neighbor (81.1%). The less accurate ML algorithms were the quadratic discriminant analysis (75.4%) and Naive Bayes (73.8%). In view of the high classification accuracy of wear, the results can be considered as very promising, and they showed the feasibility of the in-situ wear monitoring using acoustic emission.

4. Conclusions

In this study, Acoustic Emission (AE) and Machine Learning (ML) based framework has been developed to monitor and classify in-situ abrasive wear in real-time. The method has been based on quantitative wear data from real-time surface analysis of the wear scar images as the ground truth. Pin-on-disc tribotests on alumina balls against steel discs have been used as a reference test. Hard alumina ball sliding against softer steel disc in dry conditions led to abrasive wear in the contact. Real-time observation of the wear scar surface was performed with a unique pin-on-disc tribometer equipped with Digital Holographic Microscope (DHM) to quantify the wear data. The DHM phase images were post-processed, and 3D reconstruction of the images allowed observing the profiles for each lap. The DHM imaging of the wear scar depth profiles allowed in the division of the friction curve into five categories based on the percentage wear above the theoretical Hertzian wear scar width. AE signals corresponding to the five categories were processed and analyzed to extract time, frequency and time-frequency domain features. Various ML algorithms have been used to classify the abrasive wear categories, and the classification accuracies for each algorithm was obtained.

A Nano AE sensor was connected to a standard AE signal acquisition Vallen system to record the AE signals obtained for the abrasive wear during the experiments. The frequency components of the AE signals for the steel and alumina combinations were in the range of 100–300 kHz. The features the three types of waveform analysis were used as input for training the classifiers. However, before training, redundant features were eliminated based on recursive feature elimination from 304 to 50 features. Various classification algorithms based on machine learning such as SVM, Neural networks, Logistic regression, Random forest, Naive Bayes, k-Nearest neighbor, quadratic discriminant analysis, and XGboost were employed to classify the wear category during the tribotest. It was found that the SVM classifies the wear categories with the highest accuracy of 84.7%. Apart from the quadratic discriminant analysis (75.4%) and the Naive Bayes (73.8%), the other ML reached high classification accuracies ranging from 81.0 to 83.9%. Regardless of the ML algorithm, the main source of classification error is between category 3 and category 4. This misclassification results from an overlap of the features between both wear categories. This was explained by the fact that category 3 and category 4 has similar wear scar widths and depths and so AE signals. To improve the classification accuracy significantly, we proposed to combine the category 3 and category 4.

The proposed method of using processed DHM image wear scar profiles as ground truth labels and the ML framework achieves a competitive accuracy for classification of the wear categories during tribometer experiments. Consequently, we can conclude that the proposed method has significant industrial potentials for in-situ and real time wear monitoring of abrasive processes in real-world industrial applications since it requires minimum modifications of commercially available industrial machines. In this work, we have proposed an effective and robust framework to monitor abrasive wear stages using the DHM and then classify the different stages based on AE signals and machine learning algorithms. However, the AE signal acquisition has been made only in case of dry conditions of abrasive wear. Furthermore, we intend to use similar AE and ML-based methods to predict various friction failures like scuffing, fretting fatigue etc.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the Empa board of directors for the Empa internal grant of the project Pro Wear.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wear.2021.203622.

Appendix

Classification results from different machine learning algorithms

As mentioned in the article, various ML classification algorithms were tested with the feature subset data and the confusion matrix showing the classification accuracy (in %) for the wear categories are shown in Table A1.
Confusion matrices for classification of wear categories using all ML classification algorithms. Table 6 is also repeated for direct comparison.

<table>
<thead>
<tr>
<th>Predicted label [%]</th>
<th>Ground truth (true label)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 1 94%</td>
<td>Category 2 94%</td>
</tr>
<tr>
<td>Category 3 8%</td>
<td>Category 4 2%</td>
</tr>
<tr>
<td>Category 5 0%</td>
<td></td>
</tr>
</tbody>
</table>

**Linear Regression (LR)**

**Naive Bayes (NB)**

**k-nearest Neighbor**

**Quadratic Discriminant Analysis (QDA)**

**Random Forest (RF)**

**Neural Networks (NN)**

**XGBoost**

**Support Vector Machine (SVM) (repetition of Table 6)**

**References**


