Characterizing Surface-damage Progression of Spur Gears with Vibration and Oil Data

Mark Walluk[1], John Tucker[2], Adrian Hood[3], Patrick Horney[4], Allen Jones[5], Wiley Matthews [6], Nenad Nenadic [7]

[1,6,7] *Rochester Institute of Technology, Rochester, NY, 14623, USA*
mrwasp@rit.edu nxnasp@rit.edu
[2] *Naval Research Laboratory, 4555 Overlook Ave SW, Washington, DC, 20375, USA*
john.e.tucker64.civ@us.navy.mil
[3] *DEVCOM Army Research Laboratory, Aberdeen Proving Ground, 21005, MD, USA*
adrian.a.hood.civ@army.mil
[4] *NAVAIR, Cherry Point, 28533, NC, USA*
patrick.r.horney.civ@us.navy.mil
[5] *NAWCAD ASG Propulsion Systems Evaluation Facility, 22229 Elmer Road, Patuxent River, 20670,MD, USA*
allen.b.jones25.civ@us.navy.mil

ABSTRACT

We present an empirical investigation of the gradual progression of surface damage in spur gears using two statistically independent data sources: oil monitoring and vibration measurements. The experimental test stand was equipped with a commercial magnetic particle filter, and a novel test process was developed to remove particles from the magnetic filter and suspend them in oil. In addition, samples were periodically drawn from the oil tank for analysis using a LaserNet Fines particle analyzer. Furthermore, image captures of selected gear surfaces after each test period provided the ground truth information on the progression of damage. Both data-driven and classical vibration-based condition indicators were computed and compared to an image-based feature that quantified the surface condition in the ground truth images with some of the indicators showing more than 80% correlation. Oil analyses did not find significant concentrations of particles in the samples drawn from the oil tank, but found relatively large numbers of particles collected from the magnetic filter.

1. INTRODUCTION

Gearbox health diagnostics continue to show benefits for operators and maintainers of mechanical assets; therefore, this field remains an area for further scientific advancement. To that end, this empirical study was concerned with the progression of gear teeth damage from initial tolerable defects to conditions that warrant replacement. The surface damage started with general wear and progressed to micropitting and pitting that are well-documented phenomena in prior works.

The early stages of damage are associated with wear and prior research suggests that mild wear at the start of gear pair operation adequately wears-in the tooth surfaces, and reduces subsequent micropitting, followed by zero or negligible wear for the rest of the gear pair life (Morales-Espejel, Rycerz, & Kadiric, 2018).

The later stages of damage progression also continue to draw the interest of both practitioners and researchers. For example, a recent empirical study of initiation and micropitting identified the critical factors in formation of micro-pits: excessive loading, gear tooth micro-geometry, surface roughness, and lubricant film thickness (Al-Tubi, Long, Zhang, & Shaw, 2015). Interesting new work on the image-based analysis of micro-pitting (Key & Kacher, 2021) and thermal mixed elastohydrodynamic lubrication models (Li & Kahraman, 2021) may provide valuable screening criteria for scuffing wear.

Recent studies found that vibration-based machine learning models outperform their classical engineered counterparts (Hood et al., 2021), the vibration analyses in this paper employed both *classical* and engineered *condition indicators* (CIs) for damage detection. Machine learning modeling approaches based on autoencoders have established themselves as a compelling methodology for predictive maintenance, especially for the first level of capability using anomaly detection. Studies using classical CIs (Lebold, McClintic, Campbell, Byington, & Maynard, 2000; Samuel & Pines, 2005; Sharma & Parey, 2016) found that the best performing clas-

sical CI was ALR, which was introduced in (Mark, Lee, Patrick, & Coker, 2010).

The recent revolution of computer vision propelled by deep neural networks (LeCun, Bengio, & Hinton, 2015) introduced new crack detection approaches such as image segmentation, based on U-Net encoder-decoder scheme (Ronneberger, Fischer, & Brox, 2015) and its variants (Siddique, Paheding, Elkin, & Devabhaktuni, 2021; Du, Cao, Liang, Chen, & Zhan, 2020), which were very effectively applied to crack detection (Cheng, Xiong, Chen, Gu, & Li, 2018; Liu, Cao, Wang, & Wang, 2019; Hsieh & Tsai, 2020) at different scales, including structural engineering (Azimi, Eslamlou, & Pekcan, 2020; Dong et al., 2021). An example of visible pitting detection and monitoring using U-net is given in (Miltenović, Rakonjac, Oarcea, Perić, & Rangelov, 2022). Image registration was proposed for monitoring surface gear failures (Onsy, Bicker, Shaw, & Fouad, 2012), and neural networks for texture analysis of wear particles (Laghari & Hassan, 2019).

The article presented here, is organized into three main sections. Section 2 describes the gears and the experimental procedure, featuring a magnetic particle filter with a novel process for removing captured particles from the filter for measurement. Section 3 describes the oil analysis results: estimated ground truth of damage from images, computing vibration-based classical CIs, and autoencoder-based data driven CIs. Optical images of gear face contact surfaces under test acted as the ground truth for vibration monitoring. Section 4 summarizes the main results and suggests future work.

2. EXPERIMENTAL SETUP

The physical experiments were performed on a custom fabricated gear testing system comprised of a 300-horse-power inline transmission dynamometer, single-stage spur gearbox, 100 kHz data acquisition system, shaft encoder, and oil delivery system. The gearbox consisted of two NASA-designed spur gears having an AGMA Class 13 tolerance, alloy 9310 composition, diametral pitch of 8, 20-degree pressure angle, and case carburization (Krantz, 2014). Accelerometers were placed on the gearbox in radial and axial positions relative to the drive shaft to acquire vibration signals associated with gear damage (Hood et al., 2021). The gears were lubricated through a forced oil circulation system with AGMA Grade 3EP synthetic oil selected, based on the pitchline velocity and operating temperature, to purposely operate in the mixed boundary lubrication regime, thereby accelerating surface-damage.

2.1. Test stand description

The test stand was built around a transmission dynamometer that used an electric motor to apply load to the gearbox and

an eddy current brake to produce the reaction torque. Spray lubrication of the gears was provided through oil impingement on the outlet of the gear mesh with a total flow rate of 0.8 L/min. The lubricant was Super Brand synthetic gear oil having a viscosity of 108 cSt at 40 ^{o}C and was actively cooled by a heat exchanger that maintained the oil temperature near 35 ^{o}C throughout the test. The oil outlet piping from the gearbox was plumbed at a constant downward slope to a reservoir having a conical bottom geometry that terminated at a drain valve port. This geometry feature was implemented to reduce locations where particles could settle out of the oil before reaching either the drain port or the downstream filter elements where oil samples were collected. The outlet port from the oil reservoir was connected to a 20-inch-long modified magnetic filter element from Eclipse Magnetics. Further downstream of the magnetic filter was a helical gear pump that led to a 3-μm rated hydraulic paper filter and the liquid-liquid heat exchanger before returning to the gearbox.

A portable USB digital microscope with a 2MP sensor and 40 to 1000× zoom captured ground truth images of the spur gear's driving side tooth surfaces as damage progressed. The microscope magnification was approximately 88×, giving a resolution of 3.2 μm/pixel. A USB-powered LED strip light provided lighting with a custom diffuser to flatten the light and a custom-built fixture, shown in Figure 1, enabled consistent positioning of the light and the microscope. The image capture process included room darkening before imaging to reduce spurious reflections due to variable outdoor light conditions. The Microsoft Camera application captured images from the camera and stored them on the laptop computer.
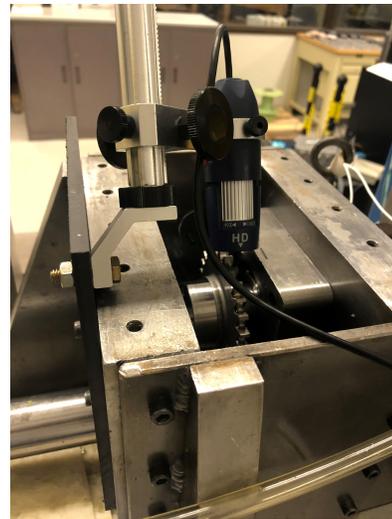

Figure 1. Setup for optical imaging of spur gear faces.

2.2. Oil sampling and magnetic filter

Oil sampling followed each test run by drawing oil at the bottom of the oil tank. Because periodic oil sampling can miss transiently generated wear particles, we installed a magnetic filter to capture particles continuously during testing. We developed a new process for removing particles collected on a sleeve around the magnet and suspending them in oil for analysis. Figure 2a shows the magnetic filter that continually collected particles as the oil circulated. Figure 2b shows the magnetic core removed from the filter, with particles forming horizontal lines at the magnetic field maxima. Figure 2c shows an example particle transferred from the magnetic core. The magnetic filter acted as a sensor that effectively integrates the wear particles as the oil circulates.

The commercial magnetic filter system was modified by installing a machined plastic spacer at the inner diameter of the filter housing to decrease the fluid channel thickness and the average distance between the magnet and particles, thus increasing the magnetic force on the ferrous gear wear debris relative to the viscous drag forces imparted by the gear oil. The preferred distance between the magnet and particles was determined by using simplified magnetic dipole models for a general understanding of the problem and by conducting benchtop experiments to study particle movement. These empirical investigations included using high-speed photography and metal 3D printing powder, having a known size distribution matching the smallest gear particles of interest, to measure the time constants associated with particle attraction to the magnet within the fluid. The associated times were used to calculate the space velocity of the oil required in the filter that ultimately defined the dimensions of the plastic spacer. The final addition to the system was a 127 - μm thick oleophobic *fluorinated ethylene propylene* (FEP) sleeve placed around the magnetic element that allowed the particles to be attracted to the magnets, which were then removed for post-test collection and analysis.

2.3. Particle Removal

During the test, as the oil circulated, magnetic forces collected particles on the FEP sleeve around the magnetic filter element. The process for removing particles from the magnet consisted of four steps. First, we folded the bottom 2 cm of the sleeve toward the top of the filter. The folding ensured that all particles remained on the sleeve. Second, we slid the sleeve upward, axially off the magnetic element. During this step, the magnet held the particles stationary until the sleeve pulled them from their position and accumulated them in the fold. The third step was cutting and removing the FEP away from the fold. In the fourth step, we unfolded the section of the FEP with particles, dunked it into a sample container with clean oil to transfer the particles, and agitated the FEP , which resulted in a visible dispersion of particles. A rare-earth bar magnet, placed on the outside wall of the container, further facilitated the transfer of the particles from FEP to the oil. After the particle extraction process, microscopy of the sleeve revealed that some particles smaller than 5 μm remained on the FEP sleeve.

LaserNet Fines wear debris analyzer processed oil samples and collected particles to determine their concentration and size distributions. Particles greater than 20 μm particles were classified into wear types. This data was used to document characteristics associated with gearbox surface-damage.

2.4. Operating conditions

Before the present experiment, we ran a 24-hour preliminary test at fixed operating conditions, with torque $\tau = 230.5$ Nm (170 ft-lb) and rotational speed $\omega = 146.6$ rad/s (1400 rpm). These operating conditions were used in earlier experiments that included the seeding of cracks (Hood et al., 2021). The torque level $\tau = 230.5$ Nm corresponded to 30% bending overload, which was also intended to accelerate surface-damage. After 24 h of testing, one of the gear teeth broke off, ending that prior test prematurely. These operating conditions served as a reference for the present experiment.

Table 1. Effect of torque on bending and contact stress safety factors

Torque [Nm] ([ft-lb])	Bending SF [%]	Contact SF [%]
230.5 (170)	69.0	75.6
162.7 (120)	97.7	90.0
149.1 (110)	106.6	94.8

Gear design equations (see, e.g., (Shigley & Mischke, 1989)) show that the fatigue stress is proportional to torque $\sigma_{fatigue} \propto \tau$. In contrast, the contact stress is proportional to the square root of torque $\sigma_{contact} \propto \sqrt{\tau}$. Thus, reducing torque was accompanied by a faster decrease in bending stress overload than contact stress overload. Table 1 shows the result of the bending safety factor for three levels of torque, where overload (OL) is related to the corresponding safety factor (SF) simply as $OL = 100 - SF$.

The progression of damage at constant torque was slow, therefore, we changed the profile of the operating conditions and introduced *agitation* by increasing torque from 169.9 Nm (125 ft-lb) to 203.4 Nm (150 ft-lb) for 20% of time, as illustrated in Figure 3. The damage progression remained slow, which motivated us to further increase the torque during the agitation period from 203.4 Nm (150 ft-lb) to 230 Nm (170 ft-lb).

Table 2 summarizes the test conditions for the test duration. The speed was not changed: it was held constant at $\omega = 146.6$ rad/s (1400 rpm) for the test duration.

(a)   (b)   (c)
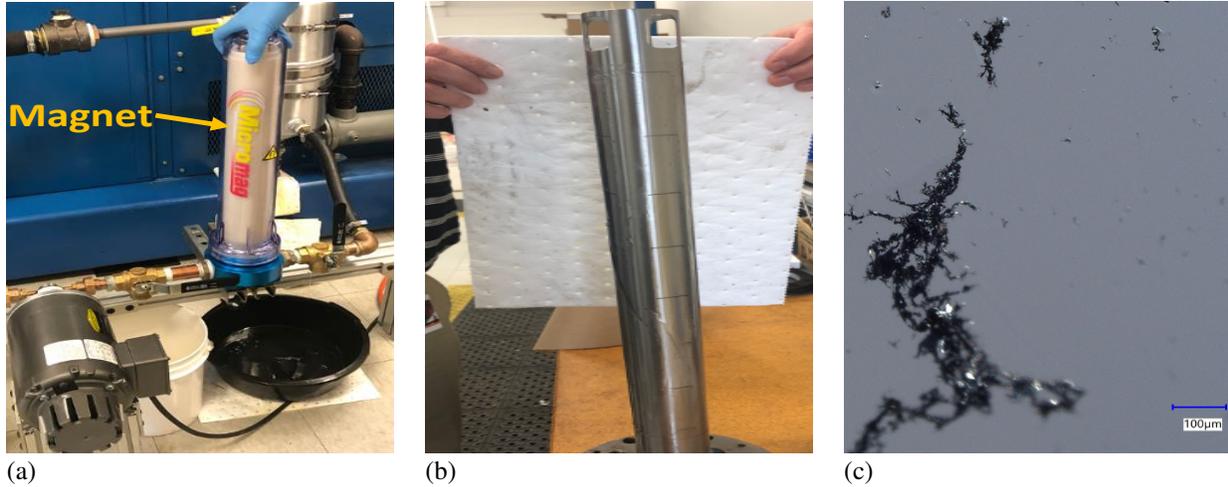
Figure 2. Magnetic particle filter: (a) in-line magnetic filter (b) magnetic element removed with captured particles forming lines at the maximum gradients of the magnetic field on FEP sleeve. (c) example of particles removed from the magnetic element.
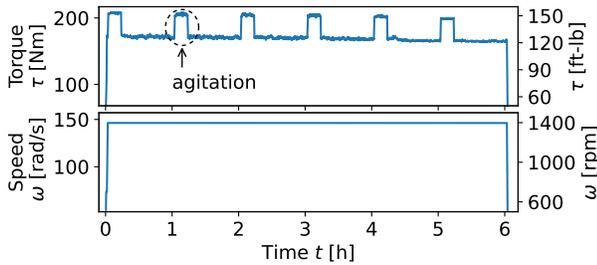

Figure 3. Operating conditions with 20% agitation

Table 2. Torque settings. The speed was held constant at ω = 146.6 rad/s (1400 rpm)

Name	Torque [Nm] ([ft-lb])	Period [h]	Dur. [h]
Fixed Torque	169.9 (125)	4	92
Agitation	169.9/203.4 (125/150)	6	258
More Agitation	169.9/230.5 (125/170)	6	188
Total			538

Table 3. Test events

Time [h]	Event	Torque [ft-lb]	Run [h]
0	Initial profile	125 ft-lb	4
92	Agitation	125/150	6
286	Removed mag. filter	125/15	6
352	Increase agitation	125/170	6
492	Re-installed mag. filter	125/170	6
526	Removed mag. filter	125/170	6
544	Test ended		

3. RESULTS AND ANALYSES

A long-term surface damage experiment provided the oil and vibration data for statistical data analyses and the development of machine learning models. Figure 4 shows a concep-
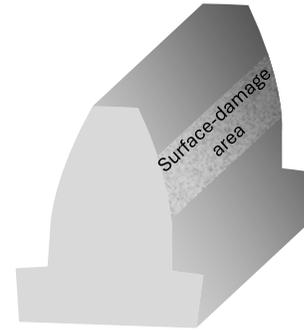

Figure 4. Surface-damage area of a single tooth (conceptual drawing)

tual view of the damaged area of a single tooth, while Figure 5 shows the image data on the damaged surface of one of the gear teeth as it progressed during the (1st or 2nd) test.

3.1. Ground truth estimation

Image-based ground truth CIs were all computed using the overall standard deviation of a grayscale image array A after an image transformation, where the image transformation consisted of two steps: histogram equalization and applying Sobel filter in the x-direction (Gonzalez & Wood, 2017). The Sobel filter was used to enhance surface brightness gradients that were associated with gear face wear.

Histogram equalization was applied to improve contrast and transform input image A, with components $a_{i,j}$, into histogram-equalized image A_e, with components $a_{e_{i,j}}$ (Kaehler & Bradski, 2016). Mathematically, the pixels
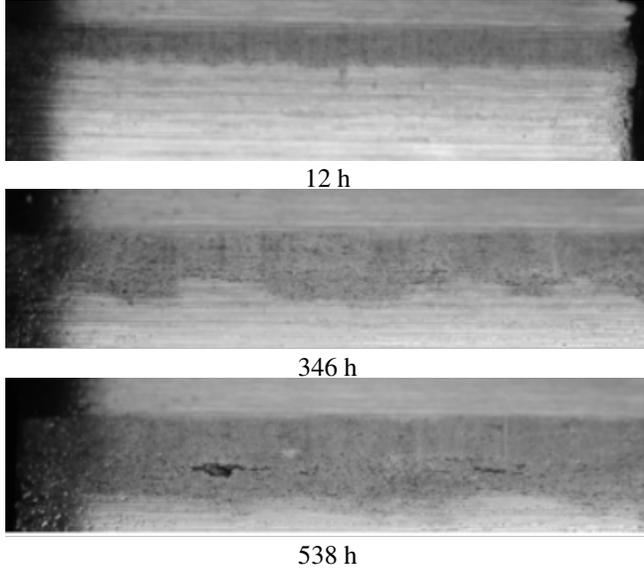
12 h

346 h

538 h

Figure 5. Progression of surface damage (a) 14 hours (b) 346 hours (c) 538 hours

intensity of the output image array are

$$a_{e_{i,j}} = \left\lfloor \frac{L-1}{NM} \sum_{n=1}^{a_{i,j}} \hat{p}[n] \right\rfloor \quad (1)$$

where L is the number of levels of the gray-scale image intensity (e.g., 256 for 8-bit pixel representation), N and M are the number of rows and columns of the image matrix $\lfloor . \rfloor$ is the floor operator (rounds the number down to the nearest integer), and $\hat{p}[n]$ is the estimated probability of the input-image pixel's intensity equal to n (the ratio of the number of pixels equal to n and the number of the elements of the matrix)

$$\hat{p}[n] = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} 1(a_{i,j} = n); \quad (2)$$

where $1(.)$ is the indicator function defined as

$$1(a_{ij} = n) = \begin{cases} 1, & a_{i,j} = n \\ 0, & a_{i,j} \neq n \end{cases} \quad (3)$$

A Sobel filter in the horizontal direction helped sharpen the image of surface damage. Filtering in the horizontal direction amplified the surface damage, whereas filtering in the vertical direction amplified the tooling-induced horizontal lines, attenuating the damage information. The filter operated on the equalized input image $\boldsymbol{A}_e$ via a convolution (denoted by $\star$). In this implementation, the filter's kernel was set to 5 and the filter is represented by a 5×5 matrix. The resulting image

A_f is thus

$$\boldsymbol{A}_f = \begin{bmatrix} -2 & -1 & 0 & +1 & +2 \\ -2 & -1 & 0 & +1 & +2 \\ -4 & -2 & 0 & +2 & +4 \\ -2 & -1 & 0 & +1 & +2 \\ -2 & -1 & 0 & +1 & +2 \end{bmatrix} \star \boldsymbol{A}_e \quad (4)$$

The corresponding OpenCV implementation of this transformation consists of two lines of code.

```
A_e = cv.equalizeHist(A)
A_f = cv.Sobel(A_e,cv.CV_64F,1,0,ksize=5)
```

Figure 6 illustrates transformed $N \times M = 500 \times 1700$ sub-images after t=406 h. A scalar ground truth estimate of surface damage was computed from the transformed image as a simple standard deviation computed for an image at the given state of the test, with one image for each of the four pre-selected teeth:

$$\sigma = \sqrt{\langle(\boldsymbol{A}_f - \langle\boldsymbol{A}_f\rangle)^2\rangle^2} = \sqrt{\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(a_{f_{i,j}} - \langle\boldsymbol{A}_f\rangle\right)^2} \quad (5)$$

where $\langle\boldsymbol{A}_f\rangle$ is the global mean of the filtered image

$$\langle\boldsymbol{A}_f\rangle = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} a_{f_{i,j}} \quad (6)$$

and $a_{f_{i,j}}$ denotes a component of the filtered image.

We computed a ground truth estimate of the surface-damage damage σ_n for each of the four teeth after each test period. Finally, the global metric was computed by averaging σ_n over the four imaged teeth

$$\langle\sigma\rangle = \frac{1}{4}\sum_{n=1}^{4} \sigma_n \quad (7)$$

Figure 7 shows how the global metric $\langle\sigma\rangle$ computed using Eq. (7) evolves from 0 to 538 h.

3.2. Vibration analysis

Several classical and data-driven CIs were computed for the long-term 538-hour test, separated into 98 runs, with periodic inspection for capturing the ground truth using optical imaging (refer to Table 2). Table 4 lists the correlations between individual vibrations-based CIs and the image-based ground truth. Of all CIs, the average likelihood of ratios (ALR) had the highest correlation coefficient with the image-based estimated ground truth ($\rho_{\langle\sigma\rangle,ALR}$= 0.84).

Classical CIs are described in detail in several review papers (e.g., (Lebold et al., 2000; Samuel & Pines, 2005; Sharma
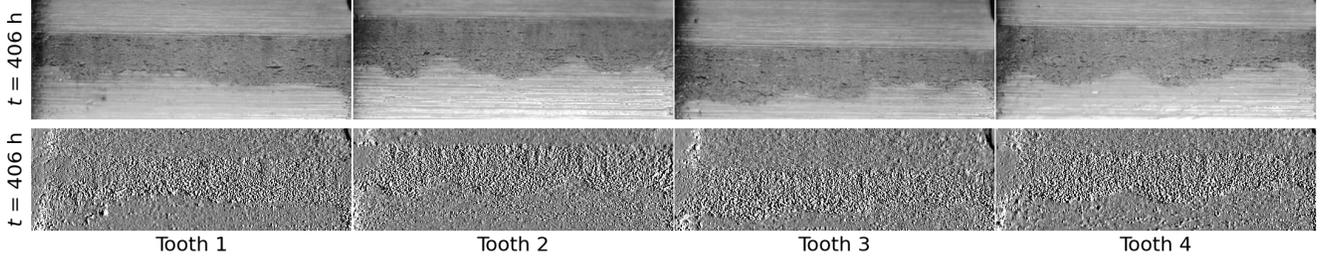
Figure 6. Four teeth imaged at $t = 406$ h: grey-scale sub-images (top row); corresponding sub-images after the transformation process (bottom row).
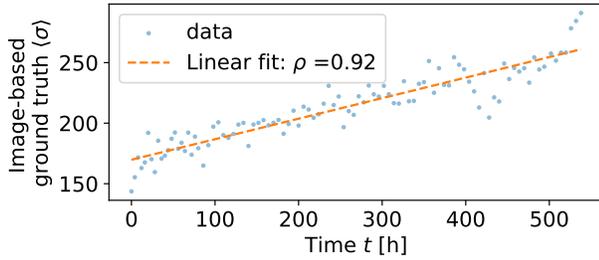

Figure 7. Evolution of the estimated image-based ground truth on surface damage.

& Parey, 2016)). These reviews do not include ALR. Since ALR is less commonly used and because it had the strongest correlation with the estimated ground truth, we describe it briefly. For more complete descriptions of this CI, refer to (Wagner, Mark, & Isaacson, 2021).

Table 4. Correlations between vibration-based condition indicators and the image-based estimated ground truth

Condition indicator	CI type	Cor. coeff. ρ [%]
ALR	deterministic*	84
Crest of energy operator	deterministic	-77
Energy ration	deterministic	-73
AE-FC-MAE	data-driven	70
Kurtosis of energy operator	deterministic	-69
Skew of energy operator	deterministic	-48
NA4	deterministic	-44
RMS	deterministic	35
FM4	deterministic	-27
M8A	deterministic	-26
M6A	deterministic	-25

The ALR CI can be computed in three steps: 1) compute the Fast Fourier Transform (FFT) of the *residual* signal denoted by $\boldsymbol{R}_b$ from the *baseline* data; 2) compute the residual of the arriving data $\boldsymbol{R}$; and 3) form the average of the logarithm of ratios

$$ALR = \frac{1}{K} \sum_{k=1}^{K} \log \frac{|r_k|}{|r_{b_k}|} \tag{8}$$

where r_k and r_{b_k} are components of $\boldsymbol{R}$ and $\boldsymbol{R}_b$, respectively.

K is the selected number of harmonics (often the total number of frequency components in the FFT).

The residual signal was obtained from the FFT of the time synchronous averaged (TSA) signal $\boldsymbol{x}_{TSA}$ by removing certain components: the first shaft and all mesh frequencies. Denoting the FFT of $\boldsymbol{x}_{TSA}$ by $\boldsymbol{X}_{TSA}$, and number of gear teeth (mesh) by N_G, residual vector $\boldsymbol{R}$ with components r_k is computed as follows

$$\begin{aligned} \boldsymbol{R} &\leftarrow \boldsymbol{X}_{TSA} \\ \boldsymbol{R}[1] &= r_1 = 0 \qquad \text{(remove shaft)} \end{aligned} \tag{9}$$

It is interesting to note that we do not need to store all baseline residuals because Eq. 8 can be rearranged to group all reference residuals together as a single scalar $\langle \log |r_{b_k}| \rangle$ as follows

$$\begin{aligned} ALR &= \frac{1}{K} \sum_{k=1}^{K} \log |r_k| - \frac{1}{K} \sum_{k=1}^{K} \log |r_{b_k}| \\ &= \frac{1}{K} \sum_{k=1}^{K} \log |r_k| - \langle \log |r_{b_k}| \rangle \end{aligned} \tag{10}$$

The TSA vector $\boldsymbol{x}_{TSA}$ is obtained from the vector of raw vibration data $\boldsymbol{a}$ and an index signal obtained from a tachometer or encoder, which enables identifying complete rotations. The process averaged acceleration over multiple shaft rotations, effectively mapping time-domain data onto angle-domain data in the range $0 \leq \theta < 2\pi$.

$$\boldsymbol{a} \mapsto \boldsymbol{x}_{TSA} \tag{11}$$

In the present case, the accelerometer vector $\boldsymbol{a}$, associated with one second of operation, had 10^5 components and was compressed into the TSA vector $\boldsymbol{x}_{TSA}(\theta)$ with the length of $K = 2^{12}$ points by averaging over 24 complete rotations. Refer to (Bechhoefer & Kingsley, 2009) for more details on TSA. Figure 8 shows ALR vs. time.

Four CIs used a subset of runs as either baseline or training data. Table 5 lists these data-driven CIs with accompanied runs used for their computations. It is important to note that the data used for CI computations were between 18 and 60
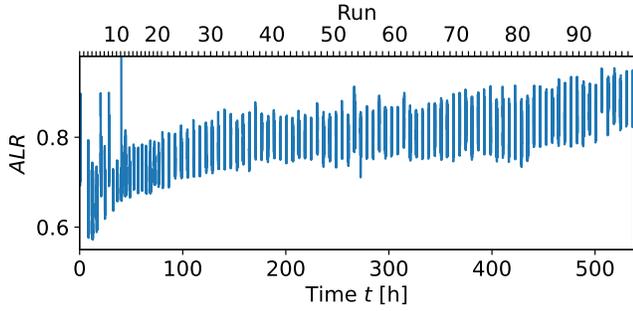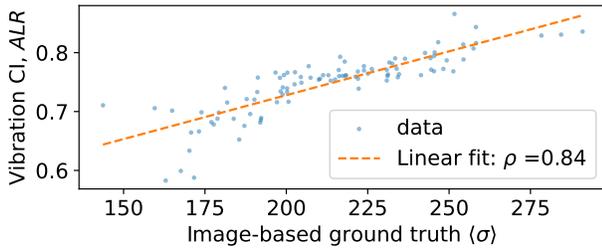

Figure 8. Evolution of ALR over time.

Figure 9. Correlation of a vibration-based condition indicator and image-based ground truth.

minutes relative to the beginning of the run because these time intervals corresponded to the steady-state data, avoiding the agitation time intervals (with higher torque), as described in Section 2.4.

For example, ALR computations used runs 1, 3, and 4 to compute the baseline residual needed for computing the actual CI, the average logarithm of ratios of residuals.

3.3. Data Driven CIs - AE-FC-MAE

Data-driven CIs, are based on the output of a trained fully-connected autoecoder. The values of these CIs were the *mean absolute error* (MAE) of autoencoder reconstruction error of the TSA features of each run, hence the name AE-FC-MAE. The MAE averages the absolute difference between the input TSA sequence, with component x_{TSA_k} and the corresponding autoencoder output $\hat{x}_{TSA_k}$.

$$CI_{AE} = \frac{1}{K} \sum_{k=1}^{K=4096} |x_{TSA_k} - \hat{x}_{TSA_k}| \qquad (12)$$

yielding one value per second.

Each of the underlying autoencoder models was symmetrical with fully-connected layers and the number of neurons from layer to layer given by 4096-64-32-16-32-64-4096, as depicted in Figure 10. The model was implemented in TensorFlow with Keras (Abadi et al., 2016) *application program-*

ming interface (Chollet, 2021; Geron, 2019). However, each model used a different selection of runs for training.

Each model used different subset of runs for training data, as shown in Table 5. The performance of the models was relatively insensitive to the subset of runs used for training, as can be observed in Table 4.

Table 5. Data-dependence of models which rely on such for either baseline or training data

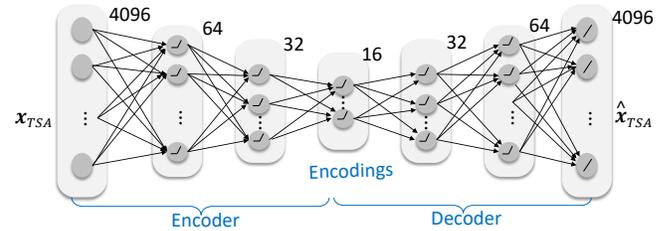| Condition indicator	Dependence type	Runs
ALR	Baseline	1,3,4
AE-FC-MAE1	Training	3,4,6,9
AE-FC-MAE2	Training	4,5,9
AE-FC-MAE3	Training	5,8,9,10

Figure 10. Autoencoder for TSA based on fully-connected layers

3.4. Oil analysis

Periodic oil samples drawn from the reservoir drain and re-suspended particles removed from the magnetic filter were analyzed using a LaserNet Fines (LNF) wear debris analyzer. The LNF analyzer examines 1-milliliter fluid volumes, and determines size and concentrations of particles in the sample that are larger than 5 μm, and classifies particles larger than 20 μm into several wear types, specifically fatigue, sliding, and cutting wear. The 3-μm system filter effectively removed particles from the fluid stream, leaving only particles that were created during the last few fluid cycles through the system. For this reason, the particle concentration in a fluid sample was related to the particle creation rate at the time the sample was drawn (Anderson & Driver, 1979).

Figure 11a shows the particle size distribution of a sample drawn at $t = 268$ h, while Figure 11b shows the total particle concentrations of the samples drawn from $t = 4$ h to $t = 388$ h. Total particle concentrations were mostly at levels below 250 per ml, and by looking at the difference between the >5 μm and >10 μm concentrations, it is seen that it was dominated by particles smaller than 10 μm in size. The fact that there were few particles greater than 10 μm was consistent with normal and pitting fatigue wear which generally creates only small particles. Particles greater than 20 μm were mostly classified as fatigue and sliding wear, with concentra-
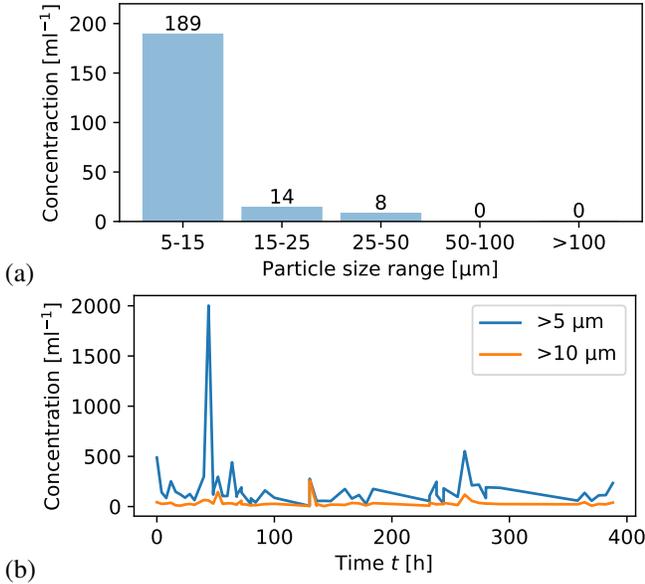
(a)

(b)

Figure 11. a) Particle size distribution for oil sample drawn at $t = 268$ h. b) Total particle concentrations during the test for sizes greater than 5 and 10 µm.

tions generally less than 3 ml^{-1} (Figure 12). There were several spikes in particle concentrations during the 388-hour test interval shown in Figures 11 and 12, which may indicate periods of increased particle generation at the time the oil sample was drawn.
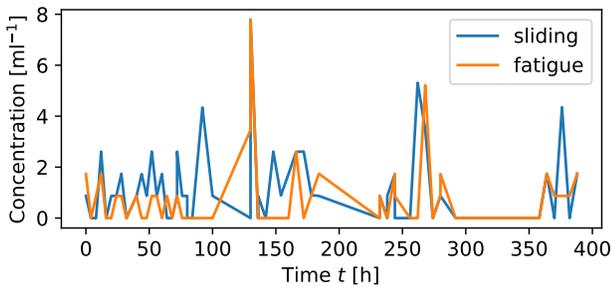

Figure 12. LNF results: sliding and fatigue particle concentrations during the test

Figure 13 shows the size distribution of particles, plotted on a logarithmic scale that were captured by the magnetic filter while it was in the oil flow from $t = 0$ to $t = 286$ h and resuspended into an oil sample. Unlike the oil samples drawn from the reservoir drain, particles were captured throughout the time the filter was in the system and acted as an integrator of the particle production rate.

This sample collected from the magnetic filter contained several hundred times more particles than ones from the reservoir and had particles larger than 100 µm in size. This sample also had similar increases in the classed particle concentrations greater than 20 µm with high concentrations of sliding and fatigue wear (Figure 14). The sliding and fatigue wear size
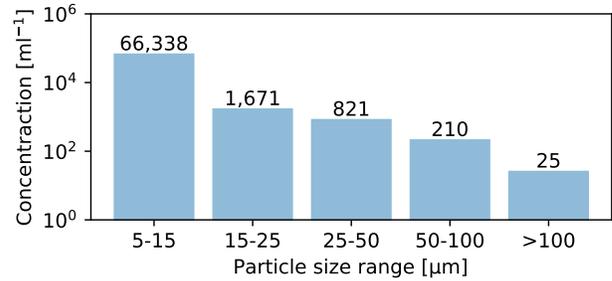

Figure 13. Logarithmic size distribution of particles removed from the magnetic filter that collected particles from t=0 to t=286 h of the test

distributions show the higher concentration of particles to be in the 25-50 µm size range, which has been seen in LNF analyses of systems experiencing machinery faults. The higher concentration in the 25-50 µm size range is due to the non-uniform size ranges in the plot and the production of larger particles due to increased surface damage.
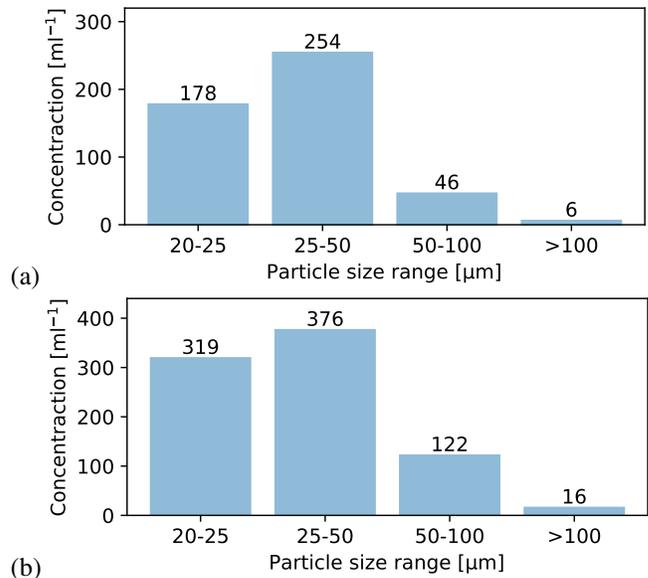

(a)

(b)

Figure 14. Wear size distributions for the magnetic filter sample (a) sliding (b) fatigue

Figure 15 shows example shapes of particles captured using the magnetic filter. The ruler at the bottom provides the scale. Several particles larger than 100 µm can be found in this image.

The much higher particle concentrations in the magnetic filter sample show the effectiveness of this device for capturing particles in the oil stream. The primary downside of the device is that there is a substantial manual effort to remove the captured particles and suspend them into a fluid sample for analysis that is not easily modified to field implementation.
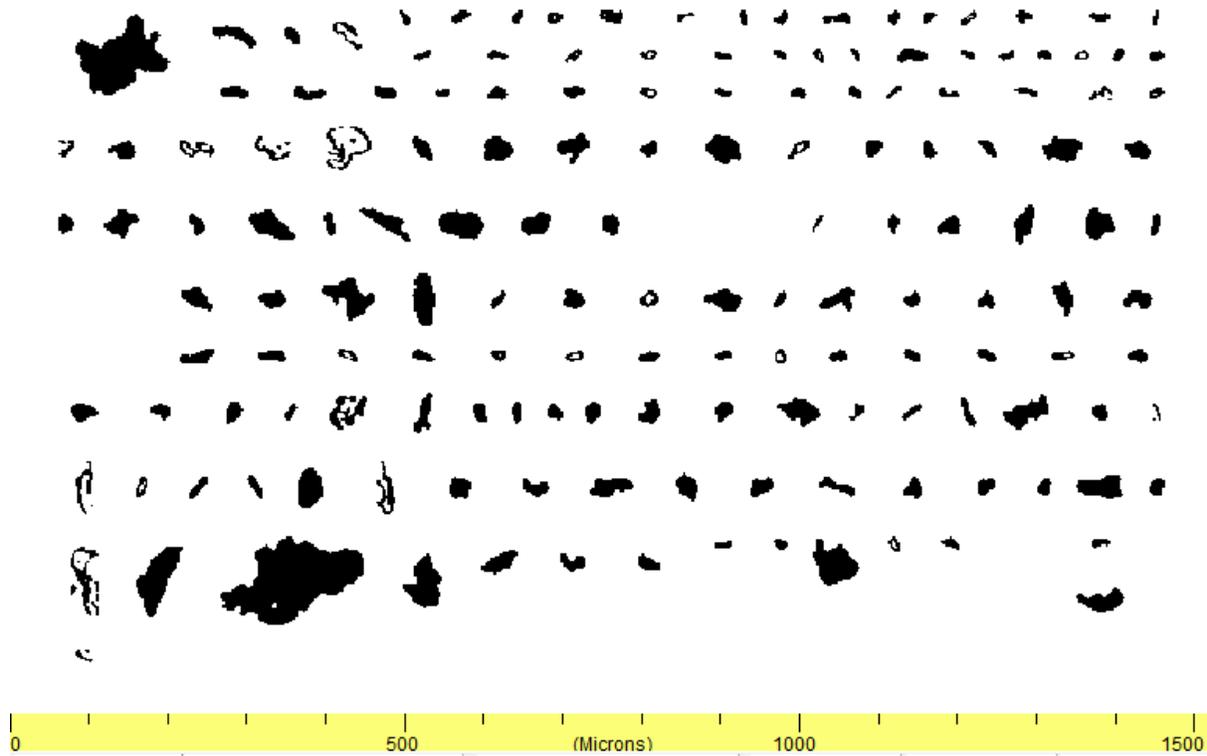
Figure 15. Oil analysis of particles captured by magnetic filter and suspended in oil

4. CONCLUSIONS AND FUTURE WORK

The study empirically examined the progression of surface damage in spur gears and included three modes of monitoring: 1) optical imaging of damaged gear surfaces; 2) characterization of particles suspended in oil and particle captured by the magnetic filter; and 3) vibration monitoring of the gearbox.

A simple, deterministic image-based feature was highly correlated with test time and vibration-based features. Ten engineered vibration CIs and three data-driven vibrations CIs were computed with four showing about 70% or better absolute correlation with the image-based condition indicator.

Particles captured by periodic oil samples after each test interval did not contain a meaningful representation of damage. However, particles captured by the magnetic filter included large particles that served as evidence of significant surface damage.

Gear failures in helicopters can be severe and lead to catastrophic outcomes, though fortunately, such failures are rare. There is a notable scarcity of publicly available datasets related to gear failures, particularly those that capture the progression of wear. It is essential to differentiate between surface damage progression and its more critical counterpart—fatigue damage. This paper, along with the accompanying dataset, aims to address this research gap by facilitating evaluation of the effectiveness of vibration-based CIs and complementary oil analysis techniques in detecting failures and accurately diagnosing their failure modes. Future work will expand both the dataset and the scope of analysis.

Specifically, future work will include inline optical and inline magnetic oil monitoring to improve data collection. Additionally, refining the image-based ground truth indicator can be improved further. For example, using the object detection and classification methods. You-Only-Look-Once (YOLO) algorithm (Redmon, Divvala, Girshick, & Farhadi, 2016), could be used to detect and classify progression of damage (micropits, pits, cracks, and spalls). The bounding boxes of detected objects will be used for development of ground-truth indicators, which can be related to vibration-based and oil CIs. Finally, feature fusion (e.g., Bayesian feature fusion) could integrate vibration and oil condition indicators, producing a more robust and reliable degradation metric.

DISCLAIMER

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Office of Naval Research.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... others (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)* (pp. 265–283).

Al-Tubi, I., Long, H., Zhang, J., & Shaw, B. (2015). Experimental and analytical study of gear micropitting initiation and propagation under varying loading conditions. *Wear*, *328*, 8–16.

Anderson, D. P., & Driver, R. D. (1979). Equilibrium particle concentration in engine oil. *Wear*, *56*(2), 415–419.

Azimi, M., Eslamlou, A. D., & Pekcan, G. (2020). Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review. *Sensors*, *20*(10), 2778.

Bechhoefer, E., & Kingsley, M. (2009). A review of time synchronous average algorithms. In *Annual conference of the phm society* (Vol. 1).

Cheng, J., Xiong, W., Chen, W., Gu, Y., & Li, Y. (2018). Pixel-level crack detection using u-net. In *Tencon 2018-2018 ieee region 10 conference* (pp. 0462–0466).

Chollet, F. (2021). *Deep learning with python*. Simon and Schuster.

Dong, C., Li, L., Yan, J., Zhang, Z., Pan, H., & Catbas, F. N. (2021). Pixel-level fatigue crack segmentation in large-scale images of steel structures using an encoder–decoder network. *Sensors*, *21*(12), 4135.

Du, G., Cao, X., Liang, J., Chen, X., & Zhan, Y. (2020). Medical image segmentation based on u-net: A review. *Journal of Imaging Science and Technology*, *64*, 1–12.

Geron, A. (2019). *Hands-on machine learning with scikit-learn, keras tensorflow* (2nd ed.). O'Reilly.

Gonzalez, R., & Wood, R. (2017). *Digital image processing* (4th ed.). Pearson.

Hood, A., Valant, C., Martuscello, J., Horney, P., Jones, A., Lantner, J., & Nenadic, N. (2021, Nov). Autoencoder-based anomaly detector for gear tooth bending fatigue cracks. *Annual Conference of the PHM Society*, *13*(1). doi: 10.36001/phmconf.2021.v13i1.3003

Hsieh, Y.-A., & Tsai, Y. J. (2020). Machine learning for crack detection: Review and model performance comparison. *Journal of Computing in Civil Engineering*, *34*(5), 04020038.

Kaehler, A., & Bradski, G. (2016). *Learning opencv 3: computer vision in c++ with the opencv library*. " O'Reilly Media, Inc.".

Key, J. W., & Kacher, J. (2021, Aug). Establishing first order correlations between pitting corrosion initiation and local microstructure in AA5083 using automated image analysis. *Materials Characterization*, *178*, 111237. doi: 10.1016/J.MATCHAR.2021.111237

Krantz, T. L. (2014). On the correlation of specific film thickness and gear pitting life. *AGMA Technical Paper*, *14FTM21*, 1-18.

Laghari, M. S., & Hassan, A. (2019). Wear particle texture analysis. In *2019 3rd international conference on imaging, signal processing and communication (icispc)* (p. 67-72). doi: 10.1109/ICISPC.2019.8935804

Lebold, M., McClintic, K., Campbell, R., Byington, C., & Maynard, K. (2000). Review of vibration analysis methods for gearbox diagnostics and prognostics. In *Proceedings of the 54th meeting of the society for machinery failure prevention technology* (Vol. 634, p. 16).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.

Li, S., & Kahraman, A. (2021). A scuffing model for spur gear contacts. *Mechanism and Machine Theory*, *156*, 104161.

Liu, Z., Cao, Y., Wang, Y., & Wang, W. (2019). Computer vision-based concrete crack detection using u-net fully convolutional networks. *Automation in Construction*, *104*, 129–139.

Mark, W. D., Lee, H., Patrick, R., & Coker, J. D. (2010). A simple frequency-domain algorithm for early detection of damaged gear teeth. *Mechanical Systems and Signal Processing*, *24*(8), 2807–2823.

Miltenović, A., Rakonjac, I., Oarcea, A., Perić, M., & Rangelov, D. (2022). Detection and monitoring of pitting progression on gear tooth flank using deep learning. *Applied Sciences*, *12*(11), 5327. Retrieved from `https://doi.org/10.3390/app12115327` doi: 10.3390/app12115327

Morales-Espejel, G., Rycerz, P., & Kadiric, A. (2018). Prediction of micropitting damage in gear teeth contacts considering the concurrent effects of surface fatigue and mild wear. *Wear*, *398*, 99–115.

Onsy, A., Bicker, R., Shaw, B., & Fouad, M. M. (2012). Application of image registration methods in monitoring the progression of surface fatigue failures in geared transmission systems. In *2012 ieee aerospace conference* (pp. 1–14).

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 779–788).

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 234–241).

Samuel, P. D., & Pines, D. J. (2005). A review of vibration-based techniques for helicopter transmission diagnostics. *Journal of sound and vibration*, *282*(1-2), 475–508.

Sharma, V., & Parey, A. (2016). A review of gear fault diagnosis using various condition indicators. *Procedia Engineering*, *144*, 253–263.

Shigley, J. E., & Mischke, C. R. (1989). Mechanical engineering design. In (Fifth ed., p. 593-595). McGraw-Hill Book Company.

Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-net and its variants for medical image segmentation: A review of theory and applications. *Ieee Access*, *9*, 82031–82057.

Wagner, M. E., Mark, W. D., & Isaacson, A. C. (2021). Implementation of the average-log-ratio alr gear-damage detection algorithm on gear-fatigue-test recordings. *Mechanical Systems and Signal Processing*, *154*, 107590.

BIOGRAPHIES

Mark Walluk received his B.S. and M.S. in Mechanical Engineering from Rensselaer Polytechnic Institute. His early career included engineering development of balance of plant components for underwater nuclear reactor and fuel cell systems, including gearboxes. Since 2007, he has been at Rochester Institute of Technology where assignments have included failure mode detection, gearbox design and testing, theoretical system analyses, prototype development, and data analytics.

Dr. John Tucker worked at the Naval Research Laboratory (NRL), Washington DC as a Research Physicist and has over 20 years experience in the development of machinery health diagnostics and the LaserNet Fines (LNF) wear debris analyzer. Dr. Tucker helped develop algorithms to translate wear particle shape to wear mechanism and then to assessment of machinery health. Dr. Tucker received a Ph. D. in Nuclear Engineering at the University of Michigan, Ann Arbor. He has published 20 papers and awarded 6 patents. In 2001, he received a Federal Laboratory Consortium Award for Technology Transfer for the LNF analyzer.

Adrian Hood received his B.S. and M.S. in Mechanical Engineering and PhD in Aerospace Engineering from the University of Maryland. He is currently a Research Mechanical Engineer at the DEVCOM US Army Research Lab in Aberdeen Proving Ground, MD. working in the Power and Propulsion Branch of the Weapons and Materials Research Directorate. His research is focused on Health and Usage Monitoring of Helicopter Transmissions, gear dynamics, tribology, and machine learning.

Patrick R. Horney received his B.S. in Aerospace Engineering from Purdue University in 2008. Since 2012 he has worked for NAVAIR in the areas of Diagnostics, Prognostics and Condidition Based Maintenance (CBM). He has established CBM/CBM+ capability for the US Navy on two aviation plaforms and serves as an advisor in the areas of PHM and CBM+.

Allen Jones received his B.S in Mechanical Engineering from the University of Maryland and his M.S. in Engineering Management from the George Washington University. He is a mechanical engineer at the Naval Air Warfare Center Aviation Division in Patuxent River, MD, working in the Air Systems Group, Propulsion and Power Engineering, Drives and Mechanical Systems Branch. His research interests include machine learning and advanced algorithms for improved diagnostics and prognostics using Health and Usage Monitoring System data

Wiley Matthews received his B.S. and M.S. in Computer Science from Rochester Institute of Technology (Rochester, NY, USA) in 2023. During his studies he completed co-ops as a Software Developer Co-op at Fujifilm North America and Machine Learning Co-op at Golisano Institute for Sustainability. He now works at Palantir Technologies as a Forward Deployed Software Engineer.

Nenad G. Nenadic received his B.S. in Electrical Engineering from University of Novi Sad (Novi Sad, Serbia) in 1996 and his MS and Ph.D. in Electrical and Computer Engineering from University of Rochester (Rochester, NY, USA) in 1998 and 2001, respectively. He joined Kionix Inc. in 2001, where he worked on development of microelectromechanical inertial sensors. Since 2005, he has been with Center for Integrated Manufacturing Studies (CIMS) at Rochester Institute of Technology, where he is currently a Research Associate Professor. His research interest include design, analysis, and monitoring of electromechanical devices and systems. He has two patents in electromechanical design. He co-authored a textbook *Electromechanics and MEMS* and is a member of IEEE.