ASME® 2020 HACKATHON

CONFERENCE
IMECE®

LOCATION
Virtual
DATES
November 14–15, 2020

2020 ASME-IMECE Hackathon: Identifying, Extracting, Analyzing of Value from Large Unstructured Data Sets in Mechanical Engineering

Virtual Event, November 14-15, 2020

In conjunction with IMECE 2020

Sponsored by

ASME Computers & Information in Engineering Division (CIE)
ASME Manufacturing Engineering Division (MED)
ASME IMECE / Advanced Manufacturing Track (AMT) &
ASME Technical Events and Content (TEC) Sector Council

ASME Manufacturing Engineering Division (MED) Centennial Celebration Event

For more details and sample datasets, please visit the Hackathon GitHub

$25 for Hackathon event (can be done as a conference add-on or stand-alone).

Click to Register for the Hackathon

Access to the Competition Sample Datasets HERE

Meeting Location: Zoom Links TBA

Important Dates:
- **Sign up Deadline November 10, 11:59 PM EDT**
- **November 14, afternoon: Hackathon kick-off**
- **November 15, afternoon: Due for Hackathon deliverables**
- **November 15, evening: Awarding ceremony**

Awards:
- **First Place: $2,000**
- **Second Place: $1,000**
- **Third Place: $500**

<u>Hackathon Problem 1</u>: Generating an Interpretable Surrogate Model for Predicting Damage Accumulation in Manufacturing

<u>Problem Statement</u>

The Bernard M. Gordon Learning Factory is a hands-on facility for engineering students, which provides modern design, prototyping, and manufacturing facilities. Many of the machines in the Learning Factory are instrumented using a sensor suite that provides monitoring capabilities. The readings from a heterogeneous set of sensors are used to report metrics continuously for many different machines. These sensors record various values every 10 minutes, such as temperature, velocity, and acceleration. The sensors also provide a computed damage accumulation measure, which is helpful for predictive maintenance.

The objective of this problem is to create an interpretable data-driven surrogate model that <u>predicts</u> the damage accumulation of machines and <u>explains</u> what has caused such damages. Specifically, you will develop a mathematical representation or equation to perform the prediction task. The technical approaches may include but are not limited, neural networks, surrogate modeling, dimensionality reduction, and even simple linear regression. However, your findings cannot merely comprise prediction results. That means, in addition to achieving satisfactory prediction accuracy, you are also expected to interpret your results, e.g., identify and explain which variables are most influential, to what extent they affect machine damage, and why they are important. The interpretation should be based on sound theories and physics underlying the observation, i.e., the damage accumulation. The resulting model and the generated insights and knowledge would be useful to the Learning Factory as a digital twin, enabling them to assess future usage scenarios for the machines and calculate the damage accumulation associated with those scenarios.

<u>Implicit Challenges</u>

Machine learning (ML) has been widely used in the Mechanical Engineering fields and has offered great potentials for improving product design, process control, and manufacturing. But **ML algorithms usually do not explain their predictions**, which is a barrier to the adoption of ML in many engineering disciplines where the underlying physics and science are critical to knowledge discovery. Recently, explainable methods for gaining an in-depth understanding of the problem-solving abilities and prediction of nonlinear ML, such as Long-Short Term Memory (LSTM) units, ensemble learning, and kernel methods, are therefore receiving increased attention.

This data and use case presents several challenges. These include:

- Is it feasible to construct a data-driven digital twin for forecasting? (Kunath et al., 2018)

- What is the best approach to identifying appropriate signals in this data with which to make predictions? (Long et al., 2019)
- How to interpret the findings based on the underlying physics?
- How to differentiate correlation vs. causation?

Datasets

Data is provided for several different machines, including three Bridgeport mills, one drill press, and one lathe. Each machine has several sensors, and each of these sensors collects data such as peak velocity, RMS velocity, peak acceleration, and temperature. A damage accumulation value is also computed from these data. Data is logged approximately every 10 minutes. Specifically, these files are provided for each machine:

1. **[machine name]week1-train.csv**
   This file contains training data for one week, including both independent variables (velocity, acceleration, etc.) and dependent variables (damage accumulation)
2. **[machine name]week2-train.csv**
   Same format as above, but for a second week.
3. **[machine name]week3-test.csv**
   This file contains data that you will use to make predictions for submission and scoring. Specifically, it contains independent variables but not dependent variables.
4. **[machine name]week3-submit.csv**
   You will use this file to submit your predictions. These files are explained in more detail under the Submission section below.

A partial example of the training data is provided below and shows one of the independent variables (peak velocity) and one of the dependent variables (damage accumulation). Every machine will have 11 independent variables and 2 dependent variables (damage accumulation in two different modes).

Machines > Lathe 1 > MIB > Y-Vertical > Peak Velocity		Machines > Lathe 1 > MIB > Y-Vertical > Damage Accumulation	
Time (UTC)	Avg(in/sec)	Time (UTC)	Avg(Damage)
2/15/20 5:08	0.0032806	2/15/20 5:08	0.9889563
2/15/20 5:18	0.0031057	2/15/20 5:18	1.0187886
2/15/20 5:28	0.0035309	2/15/20 5:28	1.012189
2/15/20 5:38	0.0023349	2/15/20 5:38	1.0183065
2/15/20 5:49	0.0040037	2/15/20 5:49	0.9838173
2/15/20 5:59	0.0019277	2/15/20 5:59	0.9948952
2/15/20 6:09	0.0037558	2/15/20 6:09	0.9767354
2/15/20 6:18	0.0027029	2/15/20 6:18	1.0100354
...	...	...	...

Submission

1. You will submit one CSV file for each machine (a total of 5 training files). Templates are provided to you and are named with the format **[machine name]week3-submit.csv**. Do not edit the time values in column 1 or headers in row 1. You should fill the remainder of columns

2 and 3 with your predictions, based on inputs from **[machine name]week3-test.csv** and the time values provided in the submission file.
2. The presentation slides. More information about the presentation and the presentation template will be introduced during the Hackathon kick-off meeting.

A partial example of a filled submission file is provided below.

Time (UTC)	Machines > Lathe 1 > MIB > X-Axial > Damage Accumulation	Machines > Lathe 1 > MIB > Y-Vertical > Damage Accumulation
2/29/20 5:06	0.86683465	0.91679842
2/29/20 5:16	0.33197709	0.3755712
2/29/20 5:26	0.07547059	0.12803265
2/29/20 5:36	0.06430619	0.04984671
2/29/20 5:46	0.41965689	0.36266437
2/29/20 5:56	0.57093488	0.18604864
2/29/20 6:07	0.38318072	0.95258113
2/29/20 6:17	0.17897555	0.48990568
2/29/20 6:26	0.18985331	0.25177106
...	...	...

Judgment Rubric

It is important to note that only 30% of your score will rely on the results from your algorithm, while the rest will be based on your approach, creativity, and presentation.

Category	Criteria	Scoring
Technical Approach (35%): Methods and algorithms of the proposed data analytics and visualization	• Requirement analysis and problem formulation • Literature review and exploration of ideas • The development and design of the idea • The readiness of the idea and the approach • The results are appropriately interpreted and can be supported by existing theories, physics, or principles. • Discovered additional (hidden) features that would be influential to machine damage beyond the provided features.	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)
Creativity and innovation (20%): A new direction in the field to approach the problem	• The technology breaks new ground • The project makes a profound break from established design • The project adds a major departure from established design • The code adds a new twist on established design • The chosen technology and design is already deeply established	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)
Results (35%) Output performance and V&V	• The objective is successfully achieved, which is measured by the Mean Squared Error and the R-squared metric over the testing set.	The best performance (10 pts) The second-best performance (7 pts)

		The third-best performance (5 pts) The top-five performance: 3 points Rest (1 pts).
Overall presentation (10%): Organization, structure, and message conveying	• Title, headings, labels: Appropriate size, location, spelling, and content • The demonstration of teamwork • Structure and Clarity • Boarder impact of the idea to ME subfields	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)

<u>Subject Matter Expert:</u>

Christopher McComb, Assistant Professor, School of Engineering Design, Technology and Professional Programs, Pennsylvania State University

Zhenghui Sha, Assistant Professor, Department of Mechanical Engineering, University of Arkansas

Binyang Song, Postdoctoral Researcher, School of Engineering Design, Technology and Professional Programs, Pennsylvania State University

Faez Ahmed, Assistant Professor, Department of Mechanical Engineering, Massachusetts Institute of Technology

<u>References:</u>

1. Kunath, M., & Winkler, H. (2018). Integrating the Digital Twin of the manufacturing system into a decision support system for improving the order management process. *Procedia CIRP*, 72, 225-231.
2. Long, Wen, Zhichen Lu, and Lingxiao Cui. “Deep learning-based feature engineering for stock price movement prediction.” *Knowledge-Based Systems* 164 (2019): 163-173.
3. Van Der Maaten, L., Postma, E. and Van den Herik, J. Dimensionality reduction: a comparative. *J Mach Learn Res* 10 (2009): 66-71.

<u>Hackathon Problem 2</u>: Smart Manufacturing – Melt-Pool Size Prediction for Powder-Bed Fusion Additive Manufacturing

<u>Problem Statement</u>

Additive manufacturing (AM) processes build parts layer-by-layer directly from 3D models. AM enables the fabrication of complex heterogeneous parts, which makes it an attractive alternative for high-value, low-volume production. However, the turnkey deployment of the technology hits consistent barriers, including poor repeatability and lack of effective part qualification tools. Optimization and control of AM processes remain a challenge. In-situ monitoring provides the capability for early detection of AM process faults and defects, enabling optimization of process parameters and closed-loop process control.

This hackathon subtopic aims to promote the use of data science in powder-bed fusion AM to predict the melt-pool size based on build commands with high accuracy. Participants are expected to train a model based on the command files and in-situ coaxial images from the first few layers to predict the melt-pool size in subsequent layers.

<u>Challenges</u>

- AM in-process data registration – how to align the build command and the coaxial images.
- How to denoise the raw image to get accurate measurements of the melt-pool size
- What factors/hidden features in the build command might affect the melt-pool formation
- How to train an effective and efficient model to achieve high prediction accuracy while minimizing computational cost
- Find an excellent way to visualize the result

<u>Datasets</u>

An experimental L-PBF build was conducted on the Additive Manufacturing Metrology Testbed (AMMT) at the National Institute of Standards and Technology (NIST). The AMMT is a fully customized metrology instrument that enables flexible control and measurement of the L-PBF process. Two cameras were installed for process monitoring, including a high-resolution camera that captures the layerwise images of the entire part and a high-speed camera used to capture melt pool images. The Galvo mirror system and the beam splitter allow the high-speed camera to focus on the current laser melting spot. Emitted light from the melt pool, through a 850 nm bandpass filter (40 nm bandwidth), is imaged on the camera sensor. On AMMT, both Galvo and laser command are updated on field-programmable gate array (FPGA) at 100 kHz. The digital commands are developed to specify the motion of the Galvo scanner of the L-PBF system. It is transformed into a time series of scanner positions and laser power as control commands.

Inconel 625 powder and build plate were used. The substrate has a dimension of 102 mm x 102 mm x 13 mm. Twelve rectangular parts (with chamfered corners) of dimensions 10 mm x 10 mm x 5 mm were laid on the substrate, with a minimum spacing of 10 mm between parts. Each part was built with a different scan strategy.

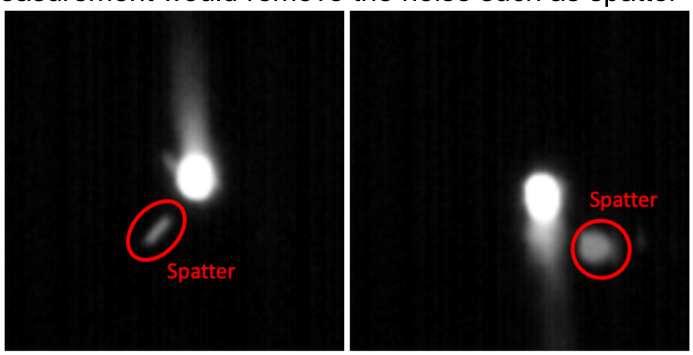
Data sets and data formats used for the study include

1) Part Design model (.stl)
2) Data sets for build command at 100KHz for all layers (.csv)
    a. Each row is the command at one step. The time interval is 10 µs.

b. The first two columns are the laser beam position represented in the substrate coordinate system.
c. The third column is the laser power.
d. The last column is the camera trigger. The camera takes one image while the trigger jumped from 0 to 2.

X (mm)	Y (mm)	Power (W)	Trigger
5.000	6.000	0	0
5.000	5.992	100	2
4.000	3.000	195	0

3) Coaxial melt-pool images for the first few layers (.bmp)
   a. Frame number follows the trigger indexing
   b. Image format: RGB, 120x120, uint8
   c. In this problem, the active melting area has grayscale >= 80
   d. The measurement would remove the noise such as spatter



Judgment Criteria

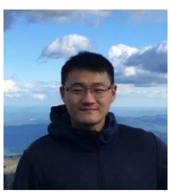
The final score will be determined by three judges based on the technical approach, results, data visualization, and presentation. Each team should submit a single column.csv file that lists predicted melt-pool size following the triggering index.

Category	Criteria	Scoring
Technical Approach (40%) Methods and algorithms of the proposed predictive model	• Requirement analysis and problem formulation • Literature review and exploration of ideas • The development and design of the idea • Scientific soundness of the approach • The creativity of the approach • The readiness of the idea and the approach • Automated workflow: data/metadata acquisition through an open interface	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)
Results (40%) Output performance and V&V	• The objective is successfully achieved, which is measured by the Mean Squared Error (MSE) over the testing layers • MSE for all regular sized melt-pool (50%) • MSE for those irregular-sized melt-pool (50%)	The best performance (10 pts) The second-best performance (7 pts) The third-best performance (5 pts)

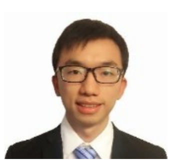
		The top-five performance: 3 points Rest (1 pts)
Data Visualization (10%) Clarity, information	• Result presentation • Data structure • Trend identification • Pattern display	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)
Overall Presentation (10%): Organization, structure, and message conveying	• Title, headings, labels: Appropriate size, location, spelling, and content • The demonstration of teamwork • Structure and Clarity • Boarder impact of the idea to ME subfields	Excellent (9-10 pts) Very good (7-8 pts) Good (5-6 pts) Limited (3-4 pts) Poor (1-2 pts)

<u>Subject Matter Expert: Mentors</u>

Yan Lu, Senior Research Scientist, Professor, System Integration Division, NIST

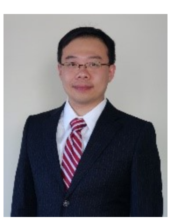


Zhuo Yang, Guest Researcher, NIST



Dehao Liu, School of Mechanical Engineering, Graduate Research Assistant, George Institute of Technology

Anh Tran, Senior Member of Technical Staff, Sandia National Laboratories



Dazhong Wu, Assistant Professor, Department of Mechanical and Aerospace engineering, University of Central Florida