Field Validation of a Universally Applicable Condition-Based Maintenance System for Mud Pumps

Dongyoung Yoon[1]*, Pradeepkumar Ashok[1], Eric van Oort[1], Pradeep Annaiyappa[2], Shungo Abe[3], and Akira Ebitani[3]

[1]The University of Texas at Austin
[2]Nabors Industries
[3]Japan Organization for Metals and Energy Security (JOGMEC)

Summary

Although mud pumps are considered to be critical rig equipment, their health monitoring currently still relies on infrequent human observation and monitoring. This approach often fails to detect pump damage at an early stage, resulting in nonproductive time (NPT) and increased well construction costs when initial damage progresses and pumps go down unexpectedly and catastrophically. Automated approaches to condition-based maintenance (CBM) of mud pumps to date have failed due to the lack of a generalized solution applicable to any pump type and/or operating conditions.

This paper presents a field-validated universally applicable solution to mud pump CBM. The system uses a sensor package that includes acoustic emission sensors and accelerometers in combination with anomaly detection deep learning data analysis to pinpoint any abnormal behavior of the pump and its components. The deep learning models are trained with undamaged normal state data only, and a damage score characterizing the extent of damage to the mud pump is calculated to identify the earliest signs of damage. The system can then generate alerts to notify the rig crew of the damage level of key mud pump components, prompting proactive maintenance actions.

Field tests were conducted while drilling an unconventional shale well in west Texas, USA, and a geothermal well in Japan (i.e., two very different drilling operations) to verify the feasibility and general applicability of the developed pump CBM solution. Sensors were attached to pump modules, and data were collected and analyzed using the deep learning models during drilling operations. During the field tests, different hyperparameters and features were compared to select the most effective ones for identifying damage while at the same time delivering low false positive rates (i.e., false alarms during normal state pump operation). The system required only several hours of normal state data for training with no prior pump information. Moreover, it correctly identified the degradation of the pump, swabs, and valves and produced early alerts several hours (in the range of 0.5–17 hours) before actual pump maintenance action was taken by the rig crew.

This generally applicable pump CBM system eliminates the environmental, health, and safety concerns that can occur during human-based observations of mud pump health and avoids unnecessary NPT associated with catastrophic pump failures. The final version of this system will be a fully self-contained magnetically attachable box containing sensors and a processor, generating simple indicators for recommending proactive pump maintenance tasks when needed.

Introduction

One of the major components in the well construction process in the oil and gas industry is the fluid circulation system, which serves many important purposes. It cleans the wellbore by removing formation cuttings, controls downhole pressure to prevent kicks and blowouts, and provides lubrication and cooling for the bit and downhole tools (Bourgoyne et al. 1986). The mud pump, which is one of the main elements of the fluid circulation system, is often regarded as the "heart" of the drilling process, because it is the machine that circulates fluid from the surface to the target depth and back (Skalle 2011).

For efficient operation of the mud pump and to avoid its unexpected failure, it is crucial for it to undergo timely maintenance and be kept in good operating condition. Fluid circulation needs to be reliable throughout the well construction operation for efficiency and safety. Mud pump damage undermines fluid circulation reliability, and unexpected failure of a mud pump may result in NPT, which can be quite costly to both the operator and the rig contractor (Pournazari et al. 2016).

The importance of monitoring the operating state and health of mud pumps has increased with the adoption of new drilling technologies. As well construction operations have become more complex and intensive, it has become necessary to add more chemicals to the fluid for improved viscosity and elevated density (Al-Yasiri and Al-Sallami 2015). In some cases, these chemicals can interact aggressively with pump parts such as elastomers (Bejger and Piasecki 2013). Moreover, deep and ultradeep reservoirs are increasingly being developed with high-pressure/high-temperature and geothermal wells, with mud pumps used in these operations facing elevated surface temperature conditions (Shaughnessy et al. 2003). This, in turn, causes more frequent pump wear and failure (Samuelson and Nirbito 2020). With more challenging drilling conditions, there can be a large difference between the predicted life of mud pumps, as specified by the manufacturer, and their actual working life in the field (Piasecki et al. 2016). Therefore, real-time mud pump monitoring in parallel with backup mud pump availability is essential to ensure a stable drilling operation that avoids unnecessary NPT and trouble costs.

*Corresponding author; email: dyyoon@utexas.edu

This paper (SPE 212564) was accepted for presentation at the SPE/IADC International Drilling Conference and Exhibition, Stavanger, Norway, 7–9 March 2023, and revised for publication. Original manuscript received for review 31 December 2023. Revised manuscript received for review 8 February 2024. Paper peer approved 14 February 2024. Supplementary materials are available in support of this paper and have been published online under Supplementary Data at https://doi.org/10.2118/212564-PA. SPE is not responsible for the content or functionality of supplementary materials supplied by the authors.

Despite the evident reasons for real-time monitoring of mud pumps, their monitoring is currently still conducted in a conventional way, relying on visual leakage monitoring or audiometric hearing by personnel for any abnormal sounds made by the pumps (Hilmawan and Basri 2021). This method is neither real-time nor continuous because observers cannot remain at the pump site throughout the entire operation. Additionally, this type of monitoring depends on the field personnel having the expertise and experience to differentiate an abnormal sound from one where all components are functioning properly. Stationing personnel near mud pumps that operate under pressures of up to 7,500 psi is also a health and safety concern (Kyllingstad and Nessjøen 2011). Moreover, the noise generated by a mud pump can exceed 100 dBA, which can potentially cause hearing damage (Teodoriu et al. 2018). These are some of the additional reasons why remote, nonhuman, real-time monitoring of mud pumps is desirable to enable CBM.

Although CBM has been applied successfully and effectively in many industries (including automotive and manufacturing industries) for a long period of time (Jardine et al. 2006; Ahmad and Kamaruddin 2012), it has to date seen only limited field application in well construction, mainly due to the high uncertainty and variability in operational conditions inherent to any drilling operation. Prior attempts to apply CBM to mud pumps have focused on tracking whether sensor data exceed a specific threshold (Kyllingstad and Nessjøen 2011; Pandurangan et al. 2019). These studies showed that failure can definitely be detected in the pumps; however, the approach is affected not only by the pump state but also by the operational conditions. A generalized, universal approach that can be applied to every pump and every possible situation has not been developed and applied to date. This difficulty in arriving at a generalized approach is complicated by the various types and manufacturers of mud pumps used in actual practice. Our goal was to develop a generalized approach that can be universally applied under different working conditions and to any type of mud pump and fully validate this approach in the field.

Methodology and Field Trial Details

CBM Approach Overview. The work described here builds on previous work reported by Yoon et al. (2022), introducing a new methodology to detect mud pump failure based on laboratory tests, with a particular focus on valve and piston damage without the need for extensive pretraining of the model. Overall, the approach involves attaching accelerometer and acoustic emission (AE) sensors to the mud pumps, collecting data, and then processing the data in real time for anomaly identification. The workflow for the system is depicted schematically in **Fig. 1.** During data preparation, the time-domain and frequency-domain features are extracted from both AE sensor and accelerometer data within 1-second intervals. These features are then synchronized based on the recorded timestamp. The pump-related data are also synchronized. Following the synchronization, the baseline is established using the auto-encoder model, which is trained based on extracted features from 6 hours of collected data. Once the baseline is constructed, continuous monitoring of pump health commences by identifying the anomalies from streaming data in real time. In addition, the baseline will be reconstructed whenever any maintenance occurs, when a new well is spudded, after mud conditioning and cementing. More details are described in the "Data Preparation" section.

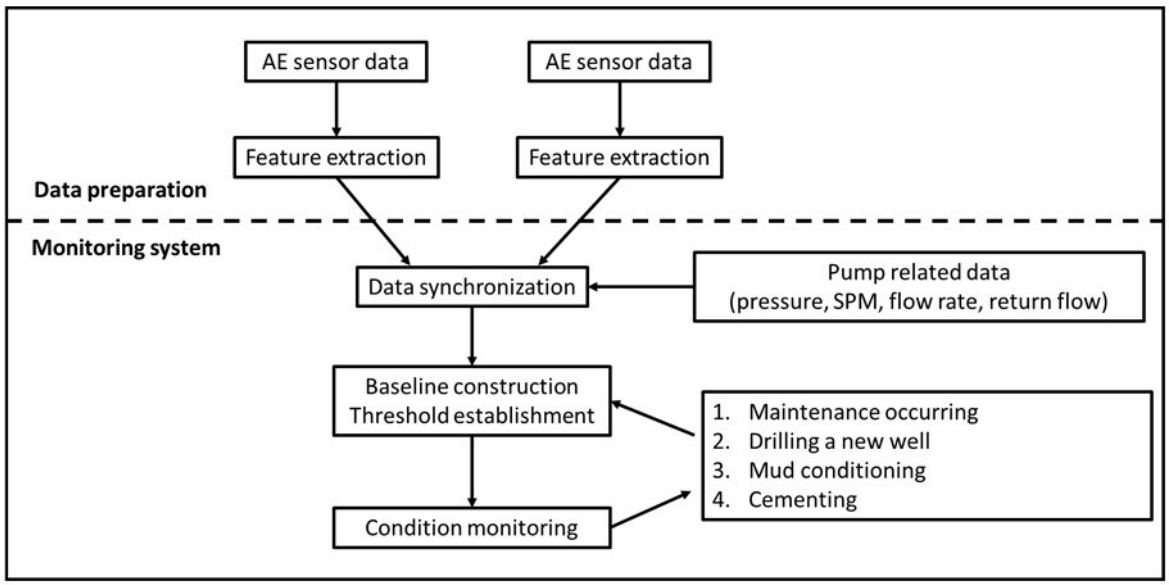

Fig. 1—Flow chart of the mud pump monitoring system's workflow.

Field Test 1: Oil and Gas Well. Field Test 1 was conducted at an onshore drilling location in Midland, Texas, and data were collected for about 10 days. Six sensor sets (each sensor set consisting of one accelerometer and one AE sensor) were attached to two mud pumps (identified as T1P2 and T1P3) operated in parallel during the drilling operation. The sensor sets were deployed near the pump suction valves. The sensor deployment and a schematic view of the test bed are shown in **Fig. 2.** During the data collection period, valve damage was observed in T1P2, and maintenance was conducted.

The operating frequency ranges for the AE sensor and accelerometers were 20–400 kHz and 0.5–4000 Hz, respectively. Due to the large volume of data associated with high sampling rates, only 1 second of data was acquired every 30 seconds.

The sampling rate used in the experiment was 8 kHz for the accelerometer and 1 MHz for the AE sensor. This rate was determined to satisfy both the minimum required sampling rate and the computational capability when considering the Nyquist theorem, by which the sampling frequency must be equal to more than two times the upper frequency of the observed signals for clear sensor data acquisition (Shannon 1949).

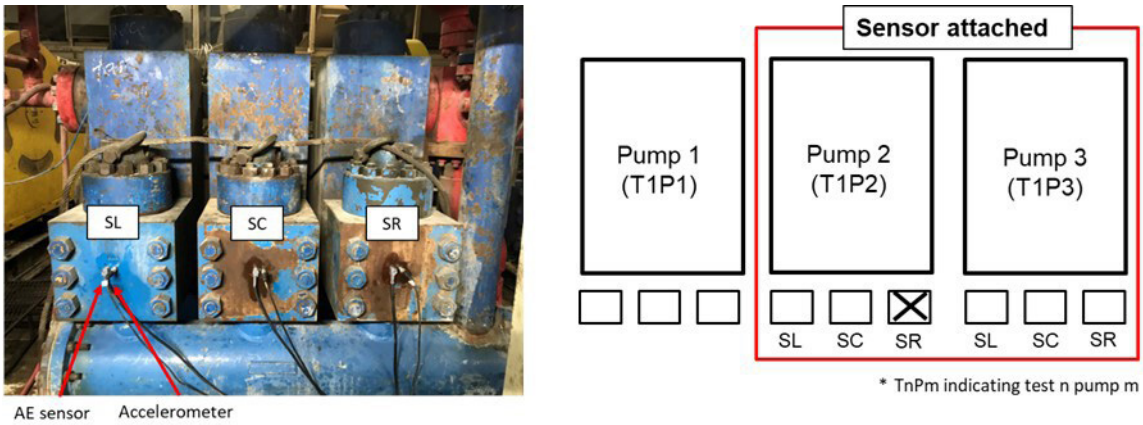

Fig. 2—(Left) Picture of the deployed sensor. (Right) Schematic view of the overall system for Field Test 1. SL, SC, and SR refer to the left, center, and right sides of the suction module, respectively.

The epoxy adhesive was used for the tight attachment of the sensors. During the trial, the sensor set at T1P2 SR was damaged because of rig operations; therefore, only data from the remaining five sensor sets was acquired and analyzed. In addition, the electronic drilling recorder data relating to the pump operation (pressure, strokes per minute, flow rate, and return flow) were analyzed to identify changes in working conditions during the drilling operation.

Field Test 2: Geothermal Well. Field Test 2 was conducted at a geothermal drilling site in Japan. The test was conducted for approximately 2 weeks, and data were collected from 12 sensor sets attached to two mud pumps. A total of six sensor sets were installed on each pump, with three sets attached to the front face of the suction modules and three to the bottom side of the discharge modules. Photos of the sensor deployment and a schematic view of the test bed are shown in **Fig. 3.**

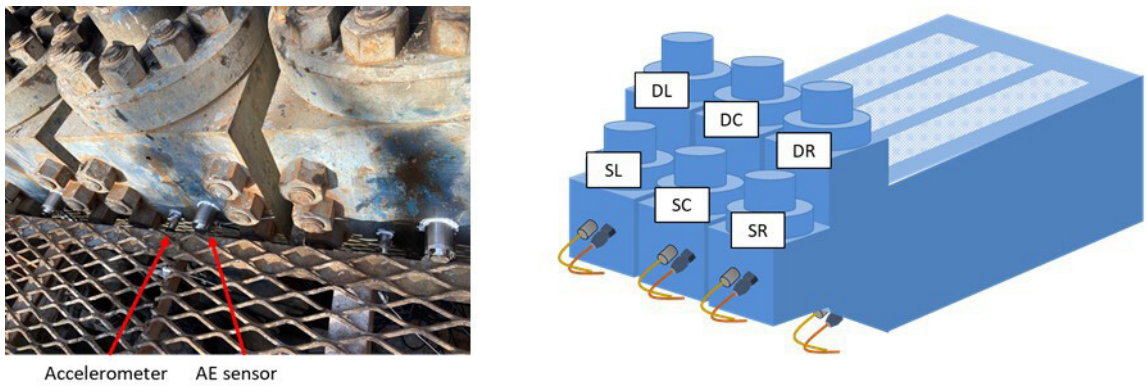

Fig. 3—Sensor deployed in Field Test 2: (left) sensor deployment on the suction modules and (right) schematic view of sensor locations.

The operating frequencies of the AE sensor and the accelerometer were in the range of 200–1200 kHz and 0.1–10 000 Hz, respectively, and the sampling rate for both the AE sensor and accelerometer was 1 MHz. The sensors were attached with magnetic holders and sealants.

In total, four swab damage events, two from each pump, were observed during the data collection phase. During maintenance, scoring marks were observed on the damaged swabs, which had caused leakage during operation.

Field Test 3: Geothermal Well. Field Test 3 took place at a geothermal drilling site in Japan. The test setup was the same as Field Test 2, where 12 sensors were attached to a total of two pumps comprising six AE sensors and six accelerometers. However, because of the issue in data collection for the accelerometer, only AE sensor data were analyzed for Field Test 3. The test was conducted for approximately 1 month, and five damage events were observed during the data collection. Specifically, three swab damages occurred in T3P1, while one valve and one swab damage were observed in T3P2. Photos of the sensor deployment and damaged swab are shown in **Fig. 4.**

Data Preparation. A total of 22 features were extracted from 1-second sensor signals of the accelerometer and the AE sensor (see Appendix A), including five time-domain features and six frequency-domain features for each sensor (Yoon et al. 2022). **Fig. 5** shows examples of raw data obtained from the T1P2 AE sensor and a 1-second time window for feature extraction.

After feature extraction, all the features and pump-related data were synchronized based on the timestamp. Nonoperational pump data were excluded by filtering out data with low AE amplitude. Also, strokes-per-minute data were considered when the data were available (T1 and T2) to filter out the data when the pump was not operating.

Fig. 4—Sensor deployed and damaged swab in Field Test 3: (left) sensor deployment on Pump 1 and (right) damaged swab from Pump 2.

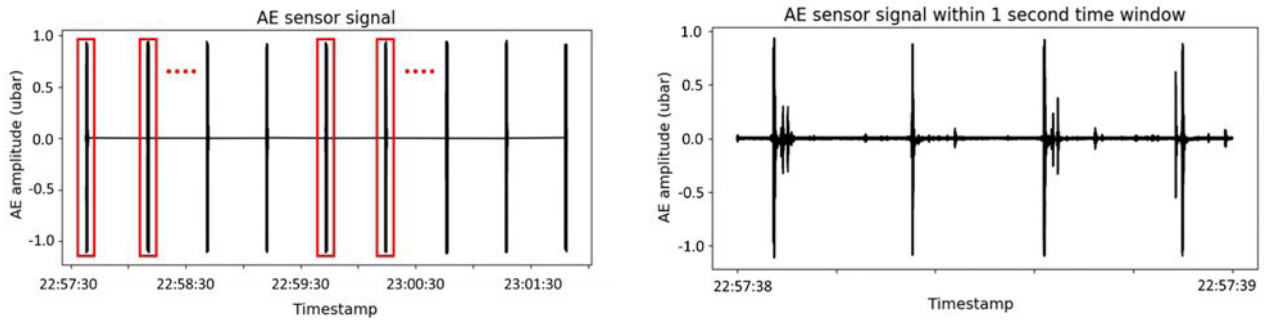

Fig. 5—Example of raw data from AE sensor located at T1P2 SL. The red boxes represent the data points within a 1-second time window that were used for feature extraction.

Sensor and Feature Selection. We systematically selected the optimal sensor setup and features across all three field tests to enhance damage detection performance while minimizing false alarms. Initially, three sensor setups were compared—accelerometer and AE sensor combined, accelerometer only, and AE sensor only. This assessment aimed to identify the most efficient setup, including the potential for system simplification. Subsequently, focusing on the selected sensor setup, feature selection was conducted using a feature reduction approach. Features contributing to false alarms were gradually eliminated by analyzing the reconstruction errors of each feature until an improvement in the overall system performance was achieved. The sensor and feature selection process is shown in **Fig. 6.**

Damage Score and Automated Alarm System. The overall workflow for the CBM system shown in **Fig. 7** consists of baseline model construction, setting of a dynamic threshold, and damage score calculation from the selected features.

The baseline operational characteristics of the pumps were constructed using an auto-encoder model (see Appendix B) using 6 hours of training data. The model's training halted automatically when the loss stopped improving for 20 epochs, thus preventing overfitting. The hyperparameters of the model consist of a learning rate of 0.05, a batch size of 16, and a LeakyReLU activation function (Maas et al. 2013). The Huber loss of the input and output data was calculated to assess the decay rate of the pump from the baseline model, as defined in Eq. B-2. A 30-minute moving average was applied to the obtained Huber loss to reduce its variations, and the resultant output was identified as the damage score.

The baseline was reconstructed every time a new well was spudded, after pump maintenance had occurred, and after mud conditioning and cementing. As an example, the analysis was conducted in three sections (identified as Time Intervals 1, 2, and 3) for T1P2 and T2P1, whereas two sections (identified as Time Intervals 1 and 2) were analyzed for T2P2, as illustrated in **Fig. 8.**

In addition, the analysis for T3P1 and T3P2 was conducted in eight sections (identified as Time Intervals 1–8), where the section division is shown in **Fig. 9.** Because the electronic drilling recorder data were not available for Field Test 3, it is plotted based on root mean square (RMS) of the AE sensor signal. The auto-encoder was retrained from 6 hours of training data during each new CBM evaluation section. For example, with T3P1 monitoring, the baseline was first constructed for 6 hours in Time Interval 1, which spans the start of data collection to the end of the drilled section. A new baseline was then constructed for 6 hours in Time Interval 2, which began after mud conditioning. The baseline was reconstructed once more in Time Interval 3 after swab replacement.

A threshold value was established, such that an alarm was generated to notify when the pump damage score exceeds the threshold. The threshold was calculated dynamically because the absolute value of the damage score is affected by the trained model. This dynamic threshold is based on the z-score and is defined as 10 standard deviations from the mean value of the damage score within the training data. Note that outliers above the 99th percentile of the damage score were filtered out before calculating the threshold. We choose this extreme value of the z-score to minimize the false alarm rate. The threshold established for T1P2-SC, Time Interval 2, is shown in **Fig. 10.**

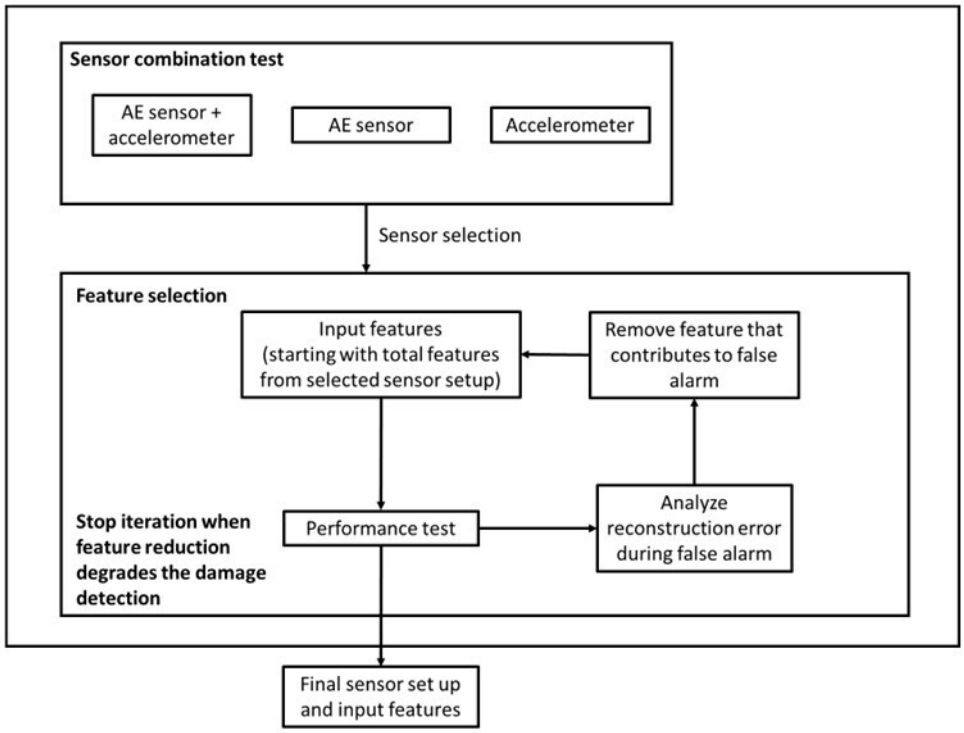

Fig. 6—Workflow of sensor and feature selection.

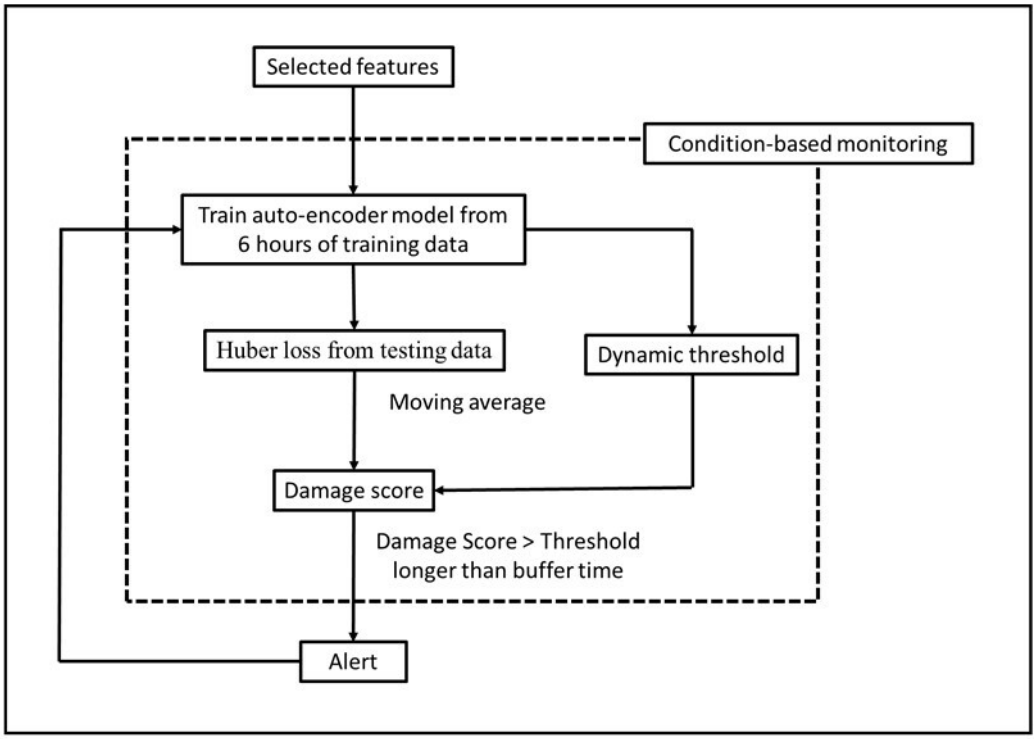

Fig. 7—Workflow of the automated alarm system for CBM.

The model also considers how long the damage score exceeds the threshold using an extra parameter termed "buffer time." This parameter is introduced to reduce false alarms from occasional spikes in the damage score. In this study, the buffer time was set to 30 minutes, which means that the system would generate alerts only if the damage score exceeded a threshold for longer than 30 minutes.

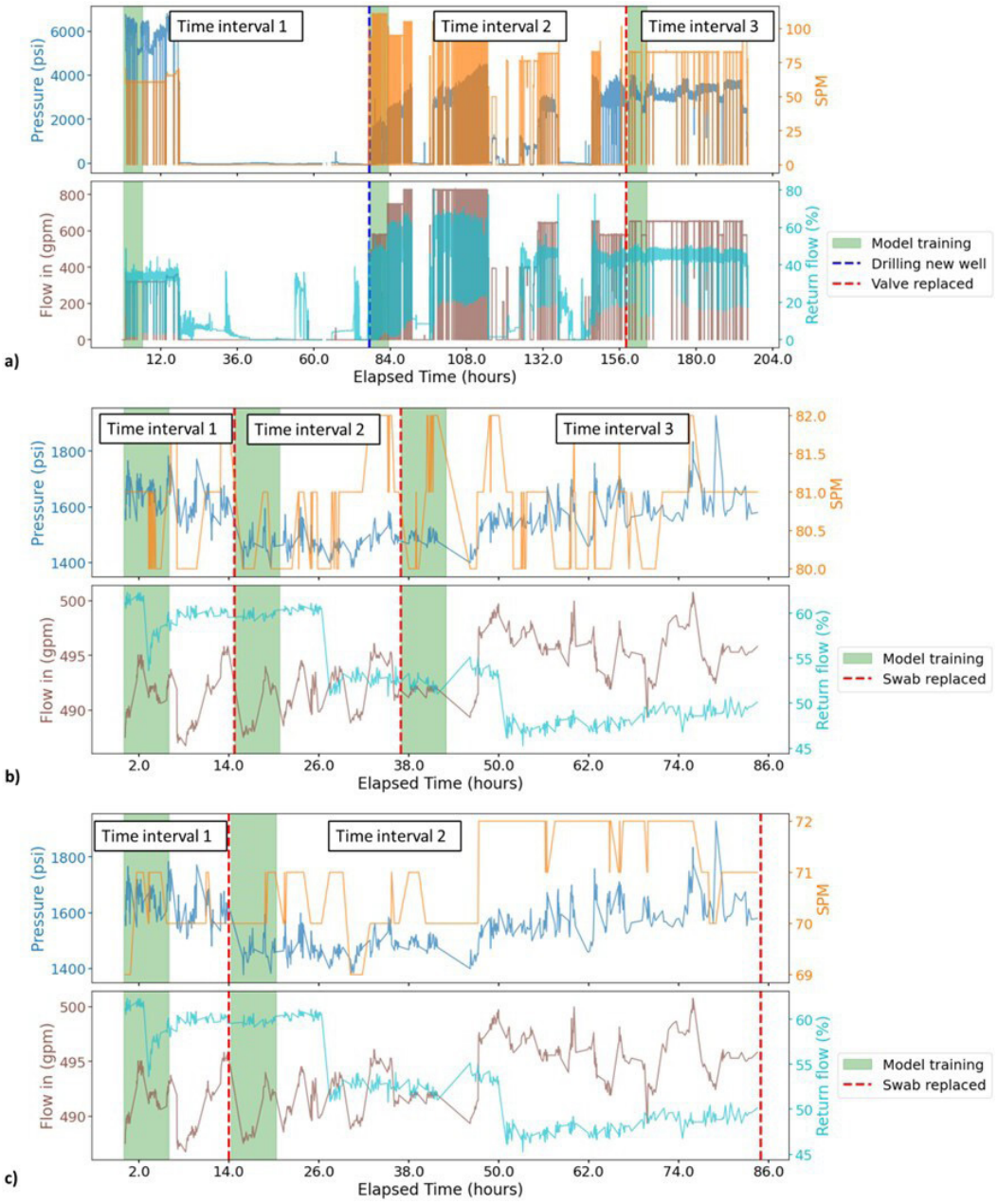

Fig. 8—Sample data illustrating the retraining of the model for Field Tests 1 and 2. The green background areas represent 6-hour training windows, and the red and blue dotted lines indicate the time when maintenance was completed and/or the new well was spudded, respectively. Data for (a) T1P2, (b) T2P1, and (c) T2P2.

Results and Discussion

In this section, monitoring results obtained from the suggested system utilizing six selected features from AE sensor data—RMS, variance (Var), spectral centroid (SC), RMS frequency (RMSF), spectral variance (SV), and spectral roll-off point (SROP)—are presented. The feature selection was conducted through the workflow shown in **Fig. 6**, where additional details can be found in Appendix C.

Damage Score and Automated Alarm System Field Test 1. The damage score with the alarm system for T1P2 Time Interval 1 is illustrated in **Fig. 11.** The system monitored the state of the pump from the damage scores calculated from each sensor set. The figure shows that there are separate damage scores for each sensor set attached to SL and SC locations (see **Fig. 2**). During Time Interval 1, the damage scores were low and well below the dynamic thresholds to prevent any alarms from being generated. This validates that the suggested system effectively avoids false alarms during the normal state operation of the pump.

Fig. 12 shows the results from T1P2 Time Interval 2. During Time Interval 2, the pump was shut off for maintenance to replace a valve. The system was able to capture this damage for about 1.5 days (equal to 17 hours of actual pump operating time) before maintenance. Note that the alert was generated from the sensors at both SL and SC locations right before the actual damage occurred.

The inclusion of a buffer time helps the system reduce the occurrence of false alarms. The damage scores at both SL and SC locations exceeded the threshold several times; however, the model did not generate an alarm because the duration for which the damage score was above the threshold was less than the buffer time (30 minutes). By contrast, the damage score for the scenario where the alarm was raised

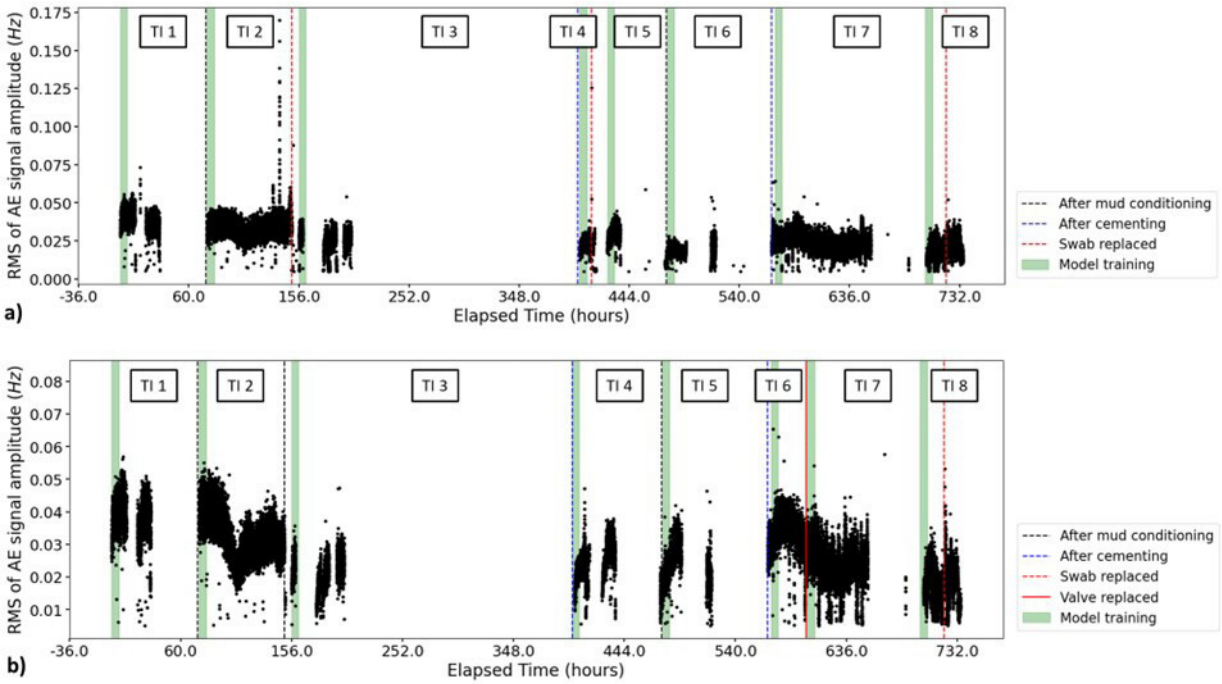

Fig. 9—Sample data illustrating the retraining of the model for Field Test 3 (a) Pump 1 (T3P1) and (b) Pump 2 (T3P2). The green background areas represent 6-hour training windows, and the red solid and dotted lines indicate the time when the valve and swab were replaced. The model is retrained after mud conditioning and cementing, which are depicted as black and blue dotted lines, respectively. Both pumps were monitored by eight different baselines throughout the entire monitoring.

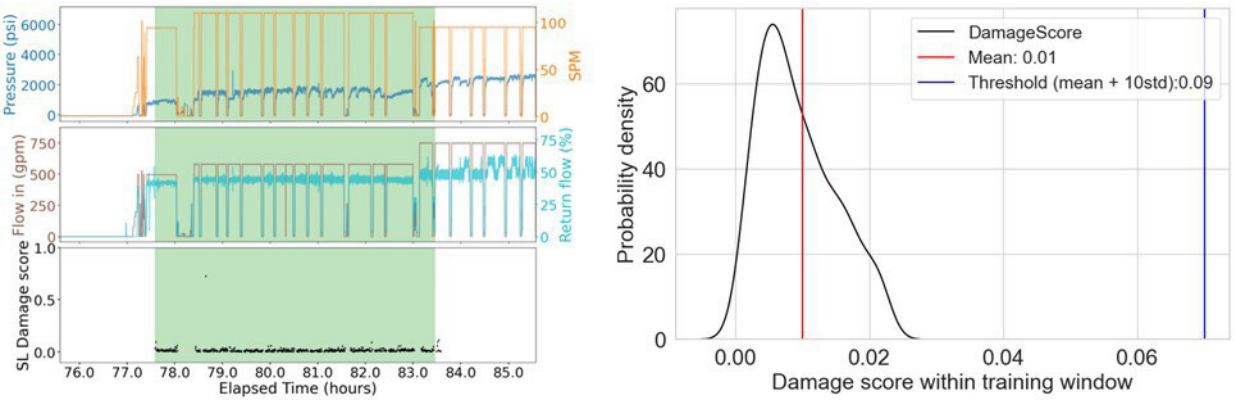

Fig. 10—Z-score-based threshold construction. (Left) Damage score within the Time Interval 2 training window for the T1P2 SL location. (Right) Probability distribution of the damage score.

before actual maintenance occurred increased continuously and remained above the threshold for a time period greater than the set buffer time.

Damage Score and Automated Alarm System: Field Test 2. In this section, the discussion is limited to the damage score from the sensors on the T2P2 Time Interval 2 suction modules for Field Test 2; the results from the remaining sensors are compiled in the Supplementary Material.

Damage to the left side of the swab was observed during Time Interval 2, with the alarm system during Time Interval 2 shown in **Fig. 13.** The system detected this damage 12 hours before maintenance. In Time Interval 2, an alert was generated from all the sensor sets (SL, SC, and SR); however, the damage score from the SL showed the highest value, correlating with the damage occurring on the left side of the pump.

Damage Score and Automated Alarm System: Field Test 3. In this section, attention is given to the damage score from the sensors on the T3P1 and T3P2 suction modules during Time Interval 2, where swab damage was observed in T3P1; the results from the remaining sensors are compiled in the Supplementary Material.

The monitoring results during Time Interval 2 for T3P1 and T3P2 are illustrated in **Fig. 14.** During the T3P1 monitoring phase, swab damage occurred on the left side of the pump, which was detected successfully by the system. The initial alarm was triggered by a spike in the damage score approximately 5 hours before the scheduled maintenance. Subsequently, additional alarms with longer durations were

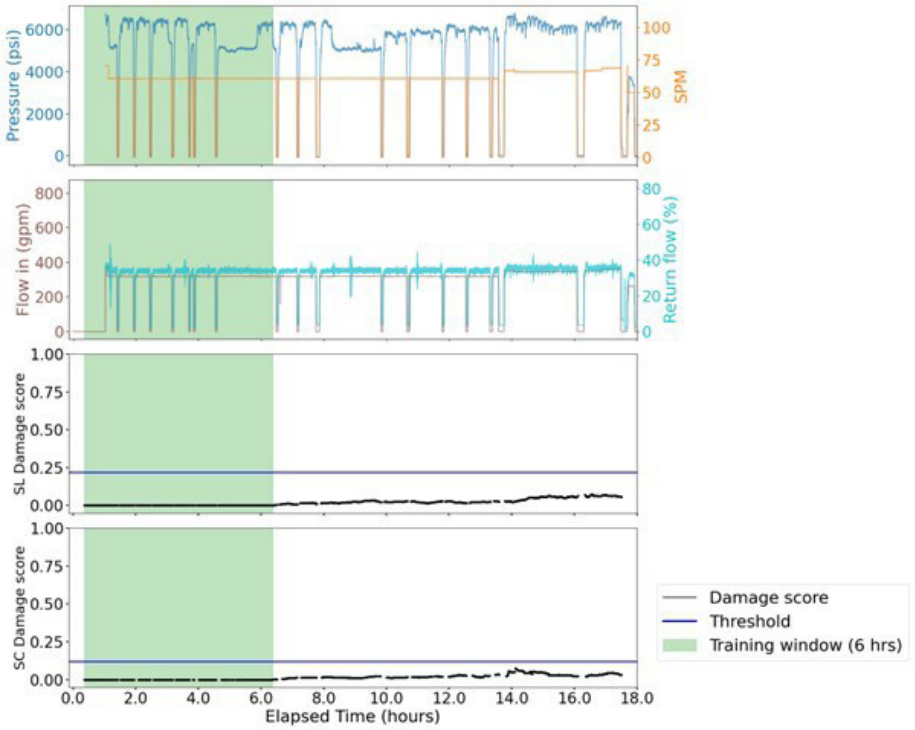

Fig. 11—Damage scores and automated alarm system for T1P2 Time Interval 1, showing baseline construction during the interval shaded in green. Zero false alarm during the normal state of the pump.

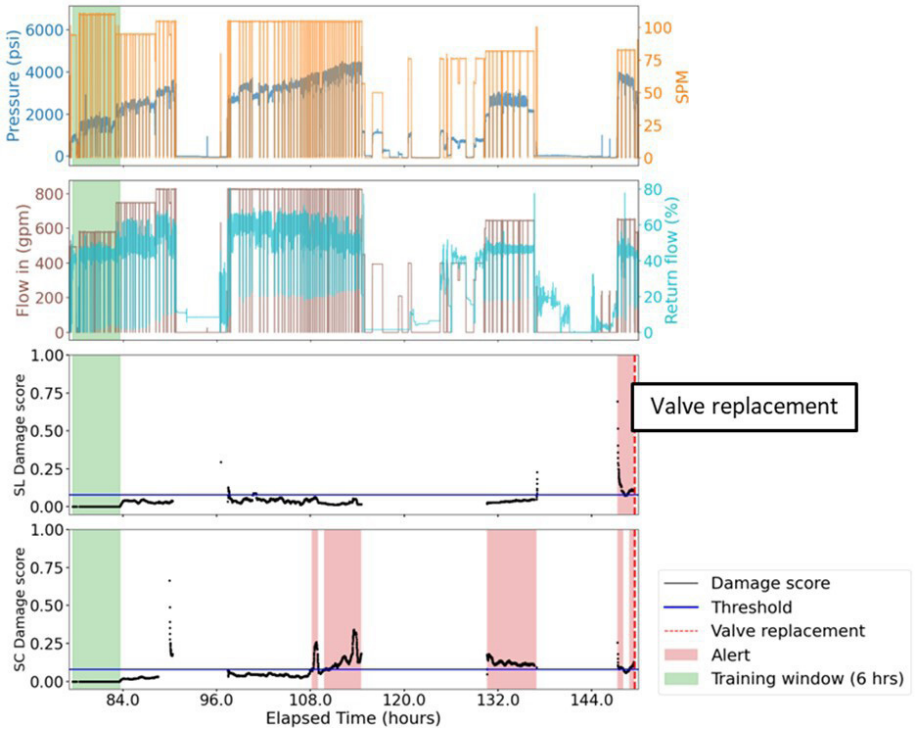

Fig. 12—Damage scores and automated alarm system for T1P2 Time Interval 2, with alerts generated in the intervals shaded in red. Valve damage was detected from the system with a forewarning time of 17 hours.

triggered about 1.5 hours before the maintenance. Specifically, the alarms were generated from the sensors located at the left and center of the suction module (SL and SC), which matches well with the actual damage location. On the other hand, despite monitoring during the same period, no alarms were generated for Pump 2 (T3P2), indicating the system effectively avoided the false alarm and only captured the signal from the damaged pump.

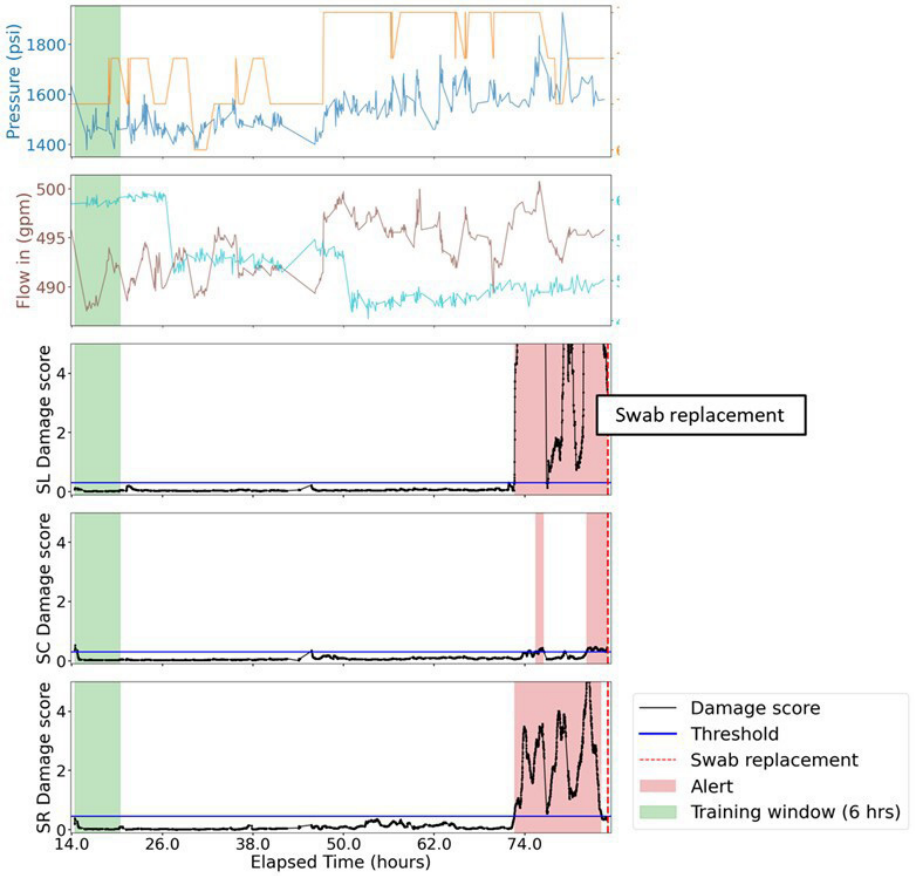

Fig. 13—Damage scores and automated alarm system for T2P2 Time Interval 2. Swab damage was detected from the system with a forewarning time of 12 hours.

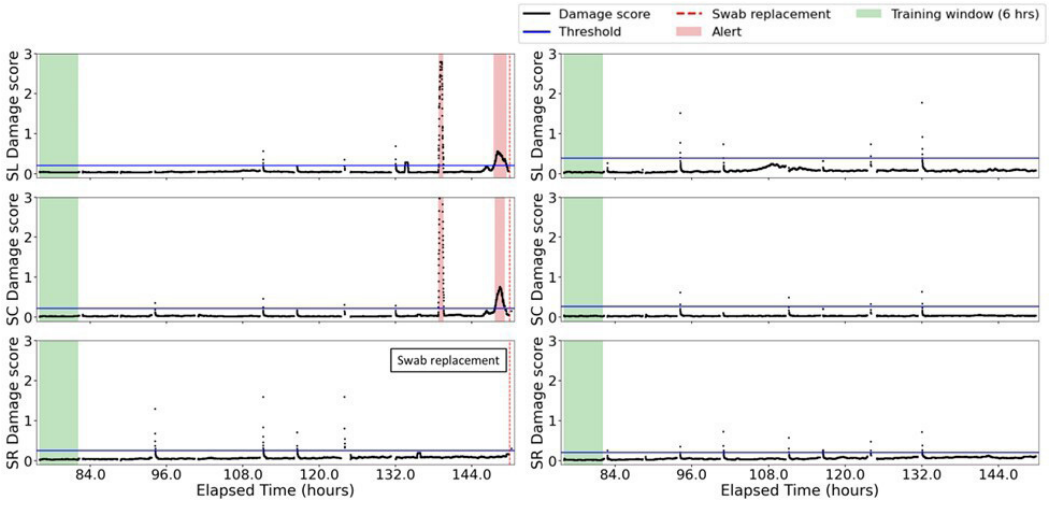

Fig. 14—Damage scores and automated alarm system for Field Test 3 Time Interval 2 from the sensor located at the suction module. (Left) Monitoring results for Pump 1 (T3P1), where the swab damage at the center suction module was successfully detected. (Right) Monitoring results for Pump 2 (T3P2), where no alerts were generated because there was no damage to Pump 2.

Damage Score and Automated Alarm System: Overall Performance. Out of 10 damage instances, including valve and swab damage, the system detected nine. The system detected damage with a minimum of 0.5 hours and a maximum of 17 hours of forewarning time. Such early-stage damage detection capability is important because late detection can result in a significant increase in damage and more severe pump failure, leading to increased repair time and associated costs. Furthermore, early detection will allow for additional time for maintenance preparation and the use of contingency measures. Furthermore, the selection of robust, universally applicable features effectively prevented false alarms, resulting in just one false alarm during each of the T2P1 and T3P1 monitoring. Also, additional findings reveal that the alarm from the system was associated with particular operations, specifically lost-circulation material (LCM), high-viscosity (Hi-vis) pill circulation that is related to lost circulation and measurement-while-drilling (MWD) surveys. These findings

also provide valuable insights for interpreting system-generated alarms. Detailed information on these specific cases is available in Appendix D. In addition, sensitivity analyses on training data size, threshold level, and sensor location are conducted and are presented in Appendix E. The overall performance of the CBM model, applying the six features from the AE sensor, is summarized in **Table 1.** If the alarm was generated before the actual maintenance, it was counted as a detected alarm. Among the generated alarms that were not related to actual damage, if there were no LCM/Hi-vis pill circulation or MWD surveys before or after the alarms, it was counted as a false alarm. The flow chart for interpreting the alarm generated from the system is depicted in **Fig. 15.**

	T1P2	T2P1	T2P2	T3P1	T3P2
Valve damage (No.)	1	0	0	0	1
Swab damage (No.)	0	2	2	3	1
Damage detection (No.)	1	1	2	3	2
Forewarning time (hours)	17	1.5	7, 12	5, 1, 0.5	0.5, 2
False alarm (No.)	0	1	0	1	0

Table 1—Summary of the diagnostic performance of the suggested model after the selection of features. The model consists of six input sizes that receive six features from AE sensor data: RMS, variance, spectral centroid, RMS of spectral centroid, spectral variance, and roll-off point.

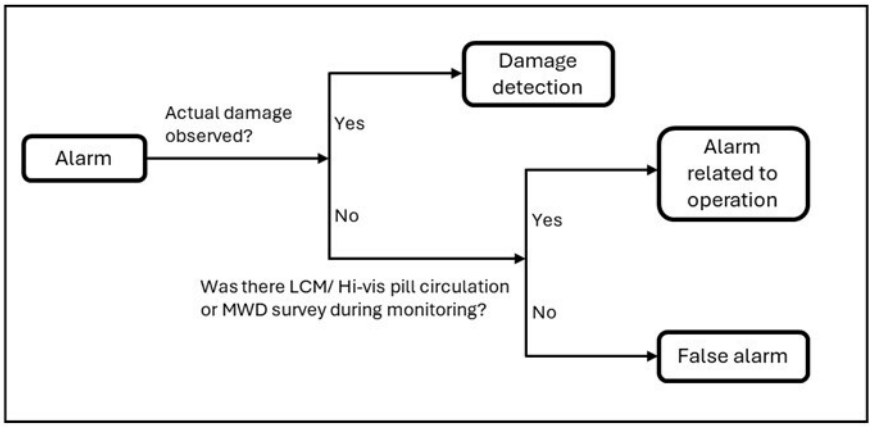

Fig. 15—Flow chart of alarm interpretation.

Conclusion

In this study, an auto-encoder-based mud pump CBM system was proposed to automatically detect valve and swab damage in mud pumps. In total, three field tests were conducted with accelerometer and AE sensor sets attached to the pumps during various drilling operations.

The model was trained using 6 hours of data collected during normal, undamaged pump operation to construct a baseline. This baseline was reconstructed whenever the operational parameters of the pump changed significantly or when maintenance occurred. Additionally, a dynamic threshold and buffer time approach for alarm generation was applied to reduce false alarms. The model successfully detected valve and swab damage with a minimum of 0.5 hours to a maximum of 17 hours of forewarning time.

The work presented here shows that early-stage damage detection can be achieved practically using the proposed method. Because the approach only requires normal state pump data for algorithm training, it offers a significant advantage in terms of real-world applicability, where very little-to-no damage data may be available to train algorithms. Also, dynamic baseline construction and threshold generation improve the system's general and universal applicability, making the model applicable under various working conditions. The proposed approach can be used as an automated real-time decision support tool for CBM of the majority of mud pumps used in the field, which is expected to dramatically reduce costs associated with operational pump failures.

Future Work

Future work on this topic will be focused on the following aspects:

1. Conducting further sensor evaluations in the field, gathering larger data sets;
2. Improving the performance of the CBM algorithm;
3. Developing a more ruggedized sensor system for field applications;
4. Extending the system to detect bearing damages, which also causes significant NPT.

Although the system detected nine damages from a total of ten actual damages, there was one undetected damage and two false alarm cases. Additional feature engineering will be conducted to improve the system's performance. In addition, different models for time-series data (i.e., using a long short-term memory (LSTM) auto-encoder) will also be tested.

A sensitivity analysis for the sampling rate will be conducted before building the ruggedized sensor system. Reducing the sampling rate is expected to significantly reduce the cost and memory requirements of the system, which will also make the system easier to ruggedize.

Acknowledgment

The authors would like to thank the Rig Automation and Performance Improvement in Drilling (RAPID) industry affiliate program at The University of Texas at Austin and its sponsors for supporting this project. Special thanks to Daryl Nygaard, Gary Miscoe, and Glen Baum for providing essential assistance in constructing and modifying the experimental setup. We also wish to thank Chandana Koritala and Michael Lockridge for their support during the yard test as well as during the field test in west Texas. We thank Atsushi Oikawa and Ryosuke Sato for their support with respect to the field test in Japan. We also thank Nabors Industries and Japan Organization for Metals and Energy Security (JOGMEC) for supporting the field tests and for permission to publish this paper.

References

Ahmad, R. and Kamaruddin, S. 2012. An Overview of Time-Based and Condition-Based Maintenance in Industrial Application. *Computers Ind Eng* **63** (1): 135–149. https://doi.org/10.1016/j.cie.2012.02.002.

Al-Yasiri, M. S. and Al-Sallami, W. T. 2015. How the Drilling Fluids Can Be Made More Efficient by Using Nanomaterials. *American J Nano Res Applications* **3** (3): 41–45. https://doi.org/10.11648/j.nano.20150303.12.

Bejger, A. and Piasecki, T. 2013. Technical Problems of Mud Pumps on Ultra Deepwater Drilling Rigs. *J of the Maritime University of Szczecin* **36** (108): 13–16.

Bourgoyne, A. T., Millheim, K. K., Chenevert, M. E. et al. 1986. *Applied Drilling Engineering*. Richardson, Texas, USA: Society of Petroleum Engineers. https://doi.org/10.2118/9781555630010.

Caesarendra, W. and Tjahjowidodo, T. 2017. A Review of Feature Extraction Methods in Vibration-Based Condition Monitoring and Its Application for Degradation Trend Estimation of Low-Speed Slew Bearing. *Machines* **5** (4): 21. https://doi.org/10.3390/machines5040021.

Helmi, H. and Forouzantabar, A. 2019. Rolling Bearing Fault Detection of Electric Motor Using Time Domain and Frequency Domain Features Extraction and ANFIS. *IET Electric Power Applications* **13** (5): 662–669. https://doi.org/10.1049/iet-epa.2018.5274.

Hilmawan, H. and Basri, H. 2021. Reducing Non-Productive Time of Mud Pump with Acoustic Emission Monitoring Techniques on Fluid End Parts. Paper presented at the 2nd ICOMERA, Malang, Indonesia, 7–9 October. https://doi.org/10.1088/1757-899X/1034/1/012066.

Jardine, A. K. S., Lin, D., and Banjevic, D. 2006. A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance. *Mechanical Systems and Signal Processing* **20** (7): 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012.

Kyllingstad, A. and Nessjøen, P. J. 2011. Condition Based Maintenance: A New Early Leak Detection System for Mud Pumps. Paper presented at the SPE/IADC Drilling Conference and Exhibition, Amsterdam, Netherlands, 1–3 March. https://doi.org/10.2118/139888-MS.

Maas, A. L., Hannun, A. Y., and Ng, A. Y. 2013. Rectifier Nonlinearities Improve Neural Network Acoustic Models. *Proceedings of the International Conferernce on Machine Learning* **30**.

Pandurangan, P., Griffith, G. E., Jaeger, T. et al. 2019. Pump Integrity Detection, Monitoring and Alarm Generation. Canadian Patent No. CA2999968C.

Piasecki, T., Bejger, A., and Kozak, M. 2016. The Assessment of the Operational Suitability of Mud Pumps on Deep Water Drillships Depends on Maintenance Strategy. *Scientific Journals of the Maritime University of Szczecin* **46** (118): 49–53. https://doi.org/10.17402/117.

Pournazari, P., Adams, D., Ashok, P. et al. 2016. Real-Time Health Monitoring of Top Drives for Offshore Operations Using Physics Based Models and New Sensor Technology. Paper presented at the SPE Deepwater Drilling and Completions Conference, Galveston, Texas, USA, 14–15 September. https://doi.org/10.2118/180356-MS.

Samuelson, A. and Nirbito, W. 2020. Problems Analysis on Preparation of Oil and Gas Drilling Rig Installation for Next Operations Readiness after HPHT (High Pressure High Temperature) Well Operation. *J. Phys.: Conf. Ser* **1519** (1): 012003. https://doi.org/10.1088/1742-6596/1519/1/012003.

Shannon, C. E. 1949. Communication in the Presence of Noise. *Proc IRE* **37** (1): 10–21. https://doi.org/10.1109/JRPROC.1949.232969.

Shaughnessy, J. M., Romo, L. A., and Soza, R. L. 2003. Problems of Ultra-Deep High-Temperature, High-Pressure Drilling. Paper presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, USA, 5–8 October. https://doi.org/10.2118/84555-MS.

Skalle, P. 2011. *Drilling Fluid Engineering*. https://bookboon.com/en/drilling-fluid-engineering-ebook?mediaType=ebook.

Teodoriu, C., Ichim, A., Srivastava, S. et al. 2018. Listening to Your Pump: Mud Pump Diagnosis and Optimization Using Audio Data. Paper presented at the 2018 AADE Fluids Technical Conference and Exhibition, Houston, Texas, USA, 10–11 April.

Yoon, D., Gul, S., Ashok, P. et al. 2022. Towards a Practical, Generally-Applicable Condition-Based Maintenance CBM System for Mud Pumps. Paper presented at the IADC/SPE International Drilling Conference and Exhibition, Galveston, Texas, USA, 8–10 March. https://doi.org/10.2118/208741-MS.

Appendix A

Feature Extraction. The time-domain features and their formulae are summarized in **Table A-1**, where x_i represents a signal for $i = 1, 2, ..., N$, and N is the number of data points. Because features were extracted from a 1-second data interval, N is the sampling rate of each sensor. In this study, N is 8,000 and 1.0×10^6 for the accelerometer and AE sensor signal, respectively, for Field Test 1, and N is 1.0×10^6 for both the accelerometer and AE sensor signal for Field Tests 2 and 3.

Frequency-domain features were extracted after converting the time-series signal into a frequency-domain representation by discrete Fourier transform given by:

$$x(k) = \sum_{n=0}^{N-1} x[n]\exp\left(-j2\pi \frac{k}{N} n\right) \text{ for } 0 \le k \le N-1, \quad \text{(A-1)}$$

where $x(k)$ is the discrete Fourier transform coefficient at frequency (in Hz), $x[n]$ is the amplitude of the nth sample in the time domain, and N is the total number of samples. In this study, the fast Fourier transform method, which is one of the most commonly used discrete Fourier transform algorithms (Jardine et al. 2006), is used.

RMS	$\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
Variance (σ)	$\frac{\sum_{i=1}^{N}\left(x_i - \frac{1}{N}\sum_{i=1}^{N} x_i\right)^2}{(N-1)}$
Crest factor (CF)	$\frac{\max\lvert x_i\rvert}{\sqrt{\frac{1}{N}\sum x_i^2}}$
Skewness (Skew)	$\frac{\sum_{i=1}^{N}\left(x_i - \frac{1}{N}\sum_{i=1}^{N} x_i\right)^3}{(N-1)\sigma^3}$
Kurtosis (Kurt)	$\frac{\sum_{i=1}^{N} F\left(x_i - \frac{1}{N}\sum_{i=1}^{N} x_i\right)^4}{(N-1)\sigma^4}$

Table A-1—Time-domain features used (Caesarendra and Tjahjowidodo 2017).

The frequency-domain features and their formulae are summarized in **Table A-2**, where $F(i)$ is the frequency for the ith frequency bin, $X(i)$ is the magnitude of the spectrum at the ith frequency bin for $i = 1, 2, ...,N$. N is the number of frequency bins; its value being half of the sampling rate and therefore 4,000 and 0.5×10^6 for the accelerometer and AE sensor signal for Field Test 1, respectively, and 0.5×10^6 for both accelerometer and AE sensor signal for Field Tests 2 and 3.

RMSF	$\sqrt{\frac{\sum_{i=1}^{N} F(i)^2 X(i)}{\sum_{i=1}^{N} X(i)}}$
SC	$\frac{\sum_{i=1}^{N} F(i) X(i)}{\sum_{i=1}^{N} X(i)}$
SV	$\frac{\sum_{i=1}^{N} (F(i) - \mathrm{SC})^2 X(i)}{N}$
Spectral skewness (SS)	$\frac{\sum_{i=1}^{N} (F(i) - \mathrm{SC})^3 X(i)}{\mathrm{SV}^{3/2} N}$
Spectral kurtosis (SK)	$\frac{\sum_{i=1}^{N} (F(i) - \mathrm{SC})^4 X(i)}{\mathrm{SV}^2 N}$
SROP	$\mathrm{SROP} = F(r)$, $where \sum_{i=1}^{r} X(i) \geq 0.85 \sum_{i=1}^{N} X(i)$

Table A-2—Frequency-domain features used (Helmi and Forouzantabar 2019).

Appendix B

Auto-Encoder. An auto-encoder is an unsupervised artificial neural network that consists of an encoder and a decoder. The encoder attempts to compress the data into a lower dimension, the latent layer, while the decoder reconstructs the original inputs from the latent layer.

Auto-encoders are widely used for anomaly detection due to their property of learning normal behavior efficiently. The model is typically trained with normal state (nonfaulty) data to capture the normal behavior during the training period. The solution can be viewed in terms of an optimization problem:

$$\min_{\boldsymbol{D},\boldsymbol{E}} L\left[X - D\left(E\left(X\right)\right)\right], \tag{B-1}$$

where X is the input data, $E(\cdot)$ is the encoder, $D(\cdot)$ is the decoder, and $L[\cdot]$ is a loss function. The idea is to train weights and biases in the model to minimize the reconstruction loss between the input and the output. Because the model is trained with only normal data, the reconstruction loss between the input data and model output will be significant when anomalous data arrive. Therefore, an anomalous condition can now be determined by setting thresholds for the loss value.

Huber loss was chosen as a loss function to reduce the model's sensitivity to the data noise, while mean squared error (MSE) is widely used as a loss function of auto-encoder mode. Huber loss combines the MSE and mean absolute error, rendering it more robust to outliers and noise compared with the MSE loss. Huber loss is quadratic for smaller errors and is linear otherwise, where the shape of the function is determined by the hyperparameter δ, where it approaches MSE when $\delta \sim 0$, and mean absolute error when $\delta \sim \infty$. The Huber loss function is expressed as

$$L_\delta(X,\hat{X}) = \begin{cases} \frac{1}{2}(X-\hat{X})^2 & \text{for } \left|X-\hat{X}\right| < \delta, \\ \delta\left|X-\hat{X}\right| - \frac{1}{2}\delta^2 & \text{otherwise.} \end{cases} \tag{B-2}$$

Appendix C

Sensor Selection: T1P2, T2P1, and T2P2. Three models were compared for T1P1, T2P1, and T2P2: Model 1 (AE sensor + accelerometer), Model 2 (accelerometer only), and Model 3 (AE sensor only). Model 1 received 22 features, while Models 2 and 3 received 11 features. The objective was to assess whether using both sensors was necessary or if one sensor sufficed. The evaluation was based on the detected damages and false alarms. False alarms were noted when alarms were triggered during the pump's normal conditions. The model depth was nine for Models 2 and 3 and 11 for Model 1 due to the higher input feature dimension.

Comparisons of the models are presented in **Tables C-1** and **C-2**. Out of five damages during Field Tests 1 and 2, Model 1 detected one instances of damage, while Model 2 and Model 3 detected three. Additionally, we observed two and three false alarms with Models 1 and 2, respectively and two false alarms with Model 3 throughout the eight monitoring time intervals. Interestingly, employing both the AE sensor and accelerometer (Model 1) resulted in minimal damage detection, which is expected to be caused by redundant features. Conversely, using either the accelerometer or the AE sensor alone enhanced model performance. The AE sensor alone effectively identified damages, indicating its potential adequacy for pump damage detection. Applying only the AE sensor would significantly simplify the system compared with utilizing both sensors. The accelerometer by itself also demonstrated promising performance. However, because only the data from the AE sensor was obtained in Field Test 3, subsequent tests and feature selection with the AE sensor data were conducted by including data from T3P1 and T3P2.

	T1P2	T2P1	T2P2
Model 1 (AE sensor + accelerometer)	0	0	1
Model 2 (accelerometer)	0	1	2
Model 3 (AE sensor)	0	1	2

Table C-1—Comparison of detected damages when using data from both the AE sensor and accelerometer, solely the accelerometer, and only the AE sensor. The total actual maintenance counts are one, two, and two for T1P2, T2P1, and T2P2, respectively.

	T1P2	T2P1	T2P2
Model 1 (AE sensor + accelerometer)	0	1	1
Model 2 (accelerometer)	1	1	1
Model 3 (AE sensor)	0	1	1

Table C-2—Comparison of number of time intervals where false alarms occurred when using data from both the AE sensor and accelerometer, solely the accelerometer, and only the AE sensor. The total time intervals with new baselines were three, three, and two for T1P2, T2P1, and T2P2, respectively.

Feature Selection. An additional test was conducted incorporating Field Test 3 (T3P1 and T3P2) and utilizing AE sensor data. The results from all field tests with Model 3 are summarized in **Table C-3**, revealing a notable number of false alarms, specifically in T3P2. Consequently, further feature selection was performed by analyzing the reconstruction error, as illustrated in the workflow in **Fig. 6**.

	T1P2	T2P1	T2P2	T3P1	T3P2
Valve damage (No.)	1	0	0	0	1
Swab damage (No.)	0	2	2	3	1
Damage detection (No.)	0	1	2	2	2
False alarm (No.)	0	1	1	2	4

Table C-3—Summary of diagnostic performance from Model 3, which applies 11 features from the AE sensor data, for valve and swab damage detection during the field application (before conducting feature selection).

For the feature selection procedure example, **Fig. C-1** illustrates the monitoring results for T3P2 during Time Interval 3, specifically from the sensor attached to the center discharge module. The figure also presents the individual contributions of AE sensor features to the overall reconstruction error. Notably, no damage was reported during this monitoring period and no specific operations were conducted, rendering the triggered alarm a false alarm. The abbreviations used are available in Appendix A for the reader's convenience. It is essential to emphasize that the primary contributor to this false alarm was identified as spectral skewness (SS) compared with other features that showed the highest reconstruction error. Correspondingly, SS was removed from the input feature, and a subsequent iteration followed. **Fig. C-2** illustrates the reconstruction error after removing the SS; the false alarm duration was reduced, but the alarm itself still exists. Therefore, the next iteration involved reducing spectral kurtosis (SK), the feature contributing most to false alarms. This iteration continued until the model demonstrated the best performance, with the highest damage detection rate and minimum false alarm occurrences. This entire feature selection process led to the retention of six features—RMS, Var, SC, RMSF, SV, and SROP—which depicted optimal performance across three field tests and the use of five pumps.

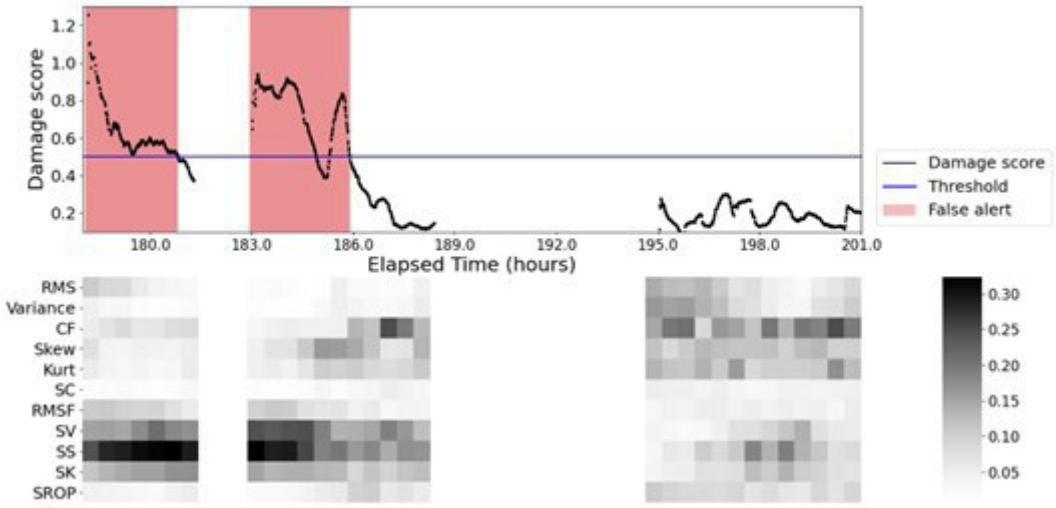

Fig. C-1—False alarm occurrences and the reconstruction error of each feature using all 11 AE sensor features for T3P2 Time Interval 3 monitoring. The false alarm persisted for approximately 8 hours from the start of monitoring. The heat map illustrates the percentage of loss attributed to each feature (reconstruction error) relative to the total loss. Importantly, spectral skewness (SS) displayed the highest contribution to false alarms during this period.

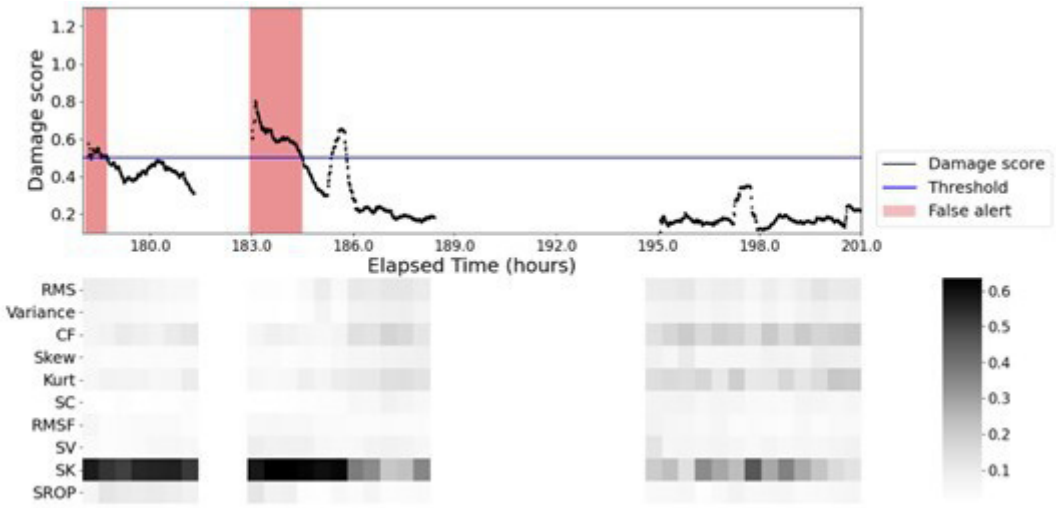

Fig. C-2—False alarm occurrences and the reconstruction error of each feature using all 10 AE sensor features after excluding spectral skewness for T3P2 Time Interval 3 monitoring. The total time for false alarms was reduced compared with the previous model with 11 features. The heat map reveals the percentage of loss (reconstruction error) attributed to each feature relative to the total loss. Markedly, spectral kurtosis (SK) displayed the highest contribution to false alarms during this period.

Appendix D

Alarms Related to Specific Operations. In this section, instances of alarm generation related to specific drilling operations are presented. Apart from the two false alarms in T2P1 (Time Interval 3) and T3P1 (Time Interval 6), where external noise was assumed to have affected the sensor, there were five additional cases where alarms were triggered despite no actual pump damage. However, these cases were not classified as false alarms, as they were caused by known reasons. All five cases were related to LCM and Hi-vis pill circulation, as well as MWD survey operations. Therefore, by merely disregarding alarms during these specific operations, operators can prevent unnecessary maintenance.

In T3P1, the first alarm emerged during Time Interval 3, as depicted in **Fig. D-1**. All sensors in the discharge module generated alarms. Remarkably, they occurred before LCM and Hi-vis pill circulation. We speculate that the circulation issue during drilling could have affected the system. This was reaffirmed by the false alarm in T3P2, again occurring before LCM circulation, as illustrated in **Fig. D-1**.

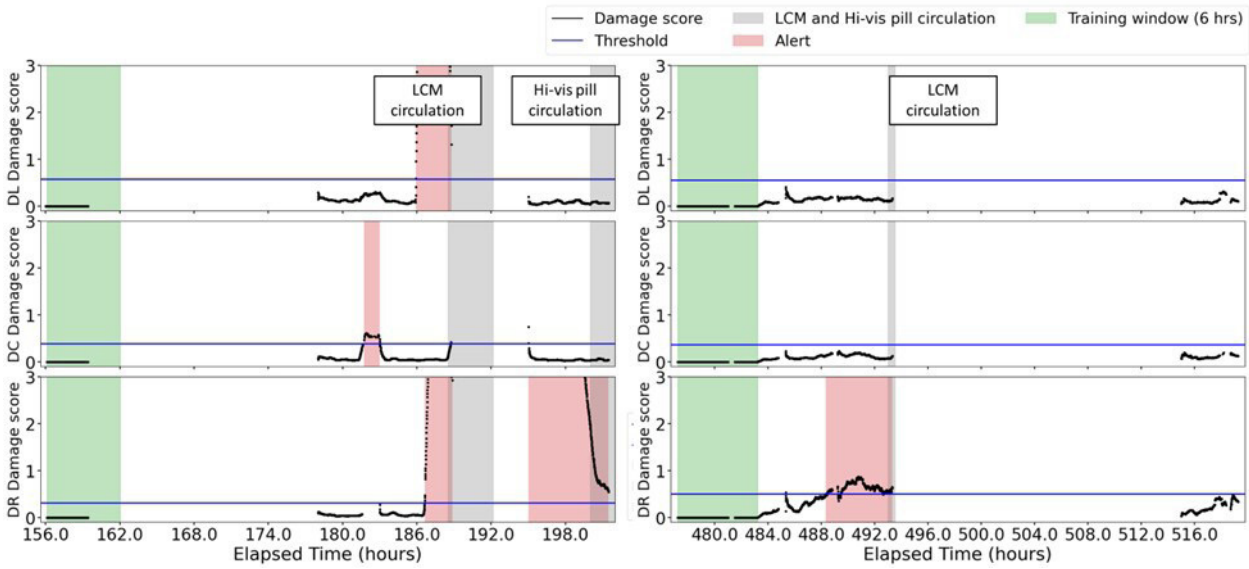

Fig. D-1—Damage scores and automated alarm system: (left) T3P1, time Interval 3 and (right) T3P2 time Interval 5. Alarms were detected within this time interval. Alarms were noted to occur before the circulation of the LCM and Hi-vis pill to the mud, suggesting that the drilling event likely influenced the sensor signals and triggered these alarms.

An additional instance of an alarm without pump damage was observed during Time Interval 7 of T3P2, as depicted in **Fig. D-2.** A series of alarms were observed throughout this monitoring period. It is essential to highlight that a consecutive MWD survey was conducted during Time Interval 7. Given that MWD surveys happened exclusively during this time frame, these MWD surveys clearly influenced the system. Specifically, the pressure pulses during the MWD survey impacted the monitoring system, as shown in **Fig. D-2**.

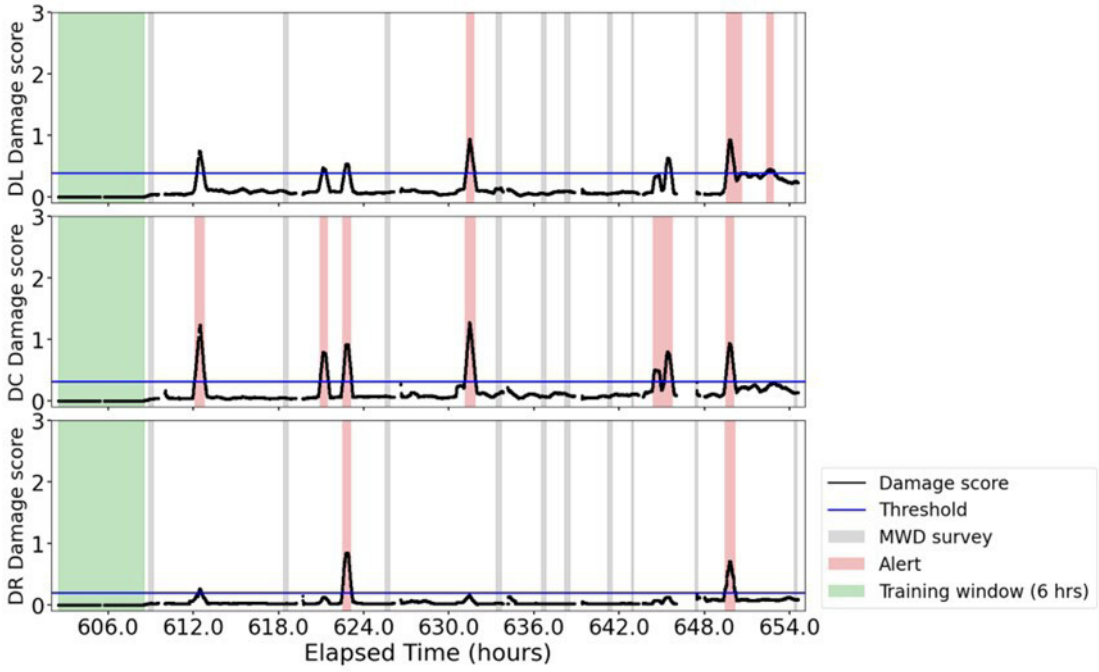

Fig. D-2—Damage scores and automated alarm system for T3P2 Time Interval 7. Alarms were detected within this time interval. During Time Interval 7, consecutive MWD surveys took place, which affected the monitoring system.

Appendix E

Effect of Training Data Size. The effect of training data size was analyzed by evaluating the performance of the models trained on different training windows. Specifically, the models were trained using four different training time windows—2 hours, 4 hours, 6 hours, and 8 hours. Because the lowest monitoring time interval in the data set was about 10 hours (Field Test 3, Pump 1, Time Interval 4), the maximum training window was set at 8 hours. The comparison of the performances is summarized in **Tables E-1 and E-2**. It can be noted that the damage detection rate is in fact affected by the training window size. When applying an excessively long training time window, there is the possibility that the training data might already contain the characteristic of a damaged state leading to a lower detection rate. It was also observed that reducing the training window results in an increased false alarm rate, as the model becomes more prone to the variance of the signal from the limited training data. Therefore, a 6-hour training window is used as the basis for the results in this paper.

	T1P2	T2P1	T2P2	T3P1	T3P2	Total
2 hours	1	1	2	3	2	9
4 hours	1	1	2	3	2	9
6 hours	1	1	2	3	2	9
8 hours	1	1	1	1	2	6

Table E-1—Comparison of detected damages when the model was trained with 2 hours, 4 hours, 6 hours, and 8 hours of training data.

	T1P2	T2P1	T2P2	T3P1	T3P2	Total
2 hours	1	1	0	1	3	6
4 hours	1	1	0	1	1	4
6 hours	0	1	0	1	0	2
8 hours	0	1	0	1	0	2

Table E-2—Comparison of false alarms when model was trained with 2 hours, 4 hours, 6 hours, and 8 hours of training data.

Effect of Threshold Level. The impact of threshold level was assessed by evaluating the performance of models with varying threshold levels. Specifically, the threshold was established at five different levels—3, 6, 8, 10, and 12 standard deviations from the mean value

of the damage score within the training data. The performance comparisons are summarized in **Tables E-3 and E-4**. No significant improvement in damage detection rate was observed by lowering the threshold level, as the damage score markedly increased during the damaged state and a model with an extreme value of 12 standard deviations threshold already detected 90% of the actual damages. However, reducing the threshold level to three standard deviations resulted in the detection of all the damages observed during the field tests. Nevertheless, it was observed that the system became more prone to false alarms with a reduced threshold level. Therefore, 10 standard deviation threshold is recommended for a balance between false alarm and damage detection rate based on the current field test data set. Despite the identical performance observed with 10 and 12 standard deviation thresholds, the decision was made to use the 10 standard deviation threshold to reduce the potential of missing a failure event in future applications.

	T1P2	T2P1	T2P2	T3P1	T3P2	Total
3 std	1	2	2	3	2	10
6 std	1	1	2	3	2	9
8 std	1	1	2	3	2	9
10 std	1	1	2	3	2	9
12 std	1	1	2	3	2	9

Table E-3—Comparison of detected damages from different threshold levels. Thresholds were established as 3, 6, 8, 10 and 12 standard deviations from the mean value of the damage score within the training data.

	T1P2	T2P1	T2P2	T3P1	T3P2	Total
3 std	2	1	1	5	3	12
6 std	1	1	1	2	2	7
8 std	1	1	0	2	1	5
10 std	0	1	0	1	0	2
12 std	0	1	0	1	0	2

Table E-4—Comparison of false alarms from different threshold levels. Thresholds were established as 3, 6, 8, 10 and 12 standard deviations from the mean value of the damage score within the training data.

Effect of Sensor Location. The sensitivity test to sensor location was conducted by comparing the system's performance from sensors located at only the suction module, only the discharge module, and both the suction and discharge modules. The comparison was conducted only with Field Tests 2 and 3 because the sensors were deployed only at the suction modules in Field Test 1. The comparisons are summarized in **Table E-5**. It was demonstrated that the sensors attached to the discharge module were more sensitive to abnormal behavior of the pump, showing both increased damage detection rate and false alarm rate compared with sensors located at suction modules. On the other hand, the monitoring results from sensors attached to the suction module presented zero false alarm rate but reduced damage detection rate. Given that swab damage was the predominant damage type during the field tests, it can be extrapolated that sensors at the discharge module were more sensitive to the damage due to their closer proximity to swab damage. Further tests are necessary to assess the impact of different sensor locations on valve damage detection rate because only one valve damage occurred during Field Tests 2 and 3. Nevertheless, based on the currently available data sets, it is recommended to attach sensors to both the suction and discharge modules to ensure comprehensive data collection. In cases where this is not feasible, attaching sensors near the discharge module is advised, with the specific goal of enhancing the damage detection rate.

	Suction Module Only	Discharge Module Only	Suction and Discharge Module
Damage detection (No.)	5	8	8
False alarm (No.)	0	2	2

Table E-5—Sensitivity to sensor locations. The total detected damages and false alarms compared for Field Tests 2 and 3.