Pipeline Integrity in the era of machine learning and artificial intelligence

PRCI-REX2021-51

Hossein Khalilpasha1, Advisian, Brisbane, Australia

Tim Edward, OneBridge Solutions, Red Deer, Canada

Abstract

The last few decades have seen significant investment in operational hardware and human resources, yet existing technology makes it difficult to keep up with the ever-increasing operational demand as assets continue to age resulting in a more complicated failure mechanism. However, in recent years with the maturity of artificial intelligence (AI) and machine learning combined with growth in the field of sensors, a shift toward data-driven and cloud solutions offer a unique opportunity for operators to flatten the curve. This abstract provides an overview of the continuous research and development conducted in the past 5 years resulting in a mature and validated solution based on machine learning and AI.

Cognitive Integrity Management ("CIM") is an advanced pipeline integrity management end-to-end SaaS application with comprehensive functionality to optimize and provide assessment planning and tracking; analyses of data integrity for regulatory compliance; dig management, real-time audit-readiness; instant business intelligence; and integration with other enterprise systems.

This study presents a case where an average pipeline system is analyzed with data gathered over a 30-year period.

To compute the financial viability of industry-standard best practices the study will provide a comparison between three different methods including pit-to-pit growth measurement done using CIM, half-life calculation, and a fixed corrosion growth rate and apply these to the same pipeline system. The study will explore the efficacy of each methodology to see which model most accurately predicts and optimizes the assessment and dig schedule for the pipeline system from both an operational and financial perspective.

Introduction

NACE International's 2020 "Cost of corrosion study" considers corrosion as the primary factor affecting the reliability of hydrocarbon pipelines throughout US. The report estimated an annual corrosion-related cost of $7 billion to monitor, replace, and maintain this infrastructure. [1]

Data plays a key role in making decisions on asset condition and developing an efficient inspection and maintenance strategy. The development of technology enables operators to collect significantly more data than previously possible.

Industry surveys [2] indicate human resource and skill shortage to remain as one of the main challenges, which means there are limited resources available to continuously monitor pipeline condition. On the other hand, the existing analysis tools and methodologies have limited application and are unable to integrate and use all available information and data. As a result, the new inspection data is not correlated effectively with historic inspection data, increasing the uncertainty, costs and also risk of failure.

Machine Learning

Machine Learning is a form of Artificial Intelligence (AI) in which a system or computer program is designed to learn on its own, potentially without any ongoing human effort or intervention. A typical approach to machine learning might be to start with a training set of labelled data in which, for a given set of inputs, the answer or result has already been determined. In that case, machine learning provides a means of working backwards from the desired output back to the input data and establishing a set of criteria or a system for deriving that desired output as closely as possible.

A practical example of this technology can be found in detecting credit card fraud where a collection of transactions, some of which were determined to be fraudulent and some of which were legitimate, are used to train a machine learning model with the goal of determining a probability of whether new transactions should be processed or denied. This is based on the system's ability to identify patterns in the input data, in some cases patterns which may not be known to any human expert that is working in the field. Rather, these patterns are derived through the brute-force process of training a model based on the labelled training dataset.

Similarly, with pipeline integrity management, tools based on Machine Learning are well suited to problems that involve "big data", where the scale of a dataset is too large for any comprehensive manual approach to be practical. Manual data analysis can often consider only a limited scale or subset of the available data and often involve filtering a dataset down to a scale where a human analyst can manage it. These modern analytical tools allow for incorporation of more of the available data into decision-making and business intelligence.

This paper presents how machine learning can be used in pipeline integrity engineering using a newly developed ML based tool.

Cognitive Integrity Management ("CIM")

Cognitive Integrity Management ("CIM") is an advanced pipeline integrity management end-to-end SaaS application. It has comprehensive functionality to optimize and provide assessment planning and tracking;

analyses of data integrity for regulatory compliance; dig management, real-time audit-readiness; instant business intelligence; and integration with other enterprise systems.

CIM provides pit-to-pit growth information for every anomaly using the full set of historical ILI data and allows linking this with the repair records dataset. The solution has three main pillars of Ingestion Algorithm, Cognitive Learning, and Business Intelligence as shown in Figure 1.

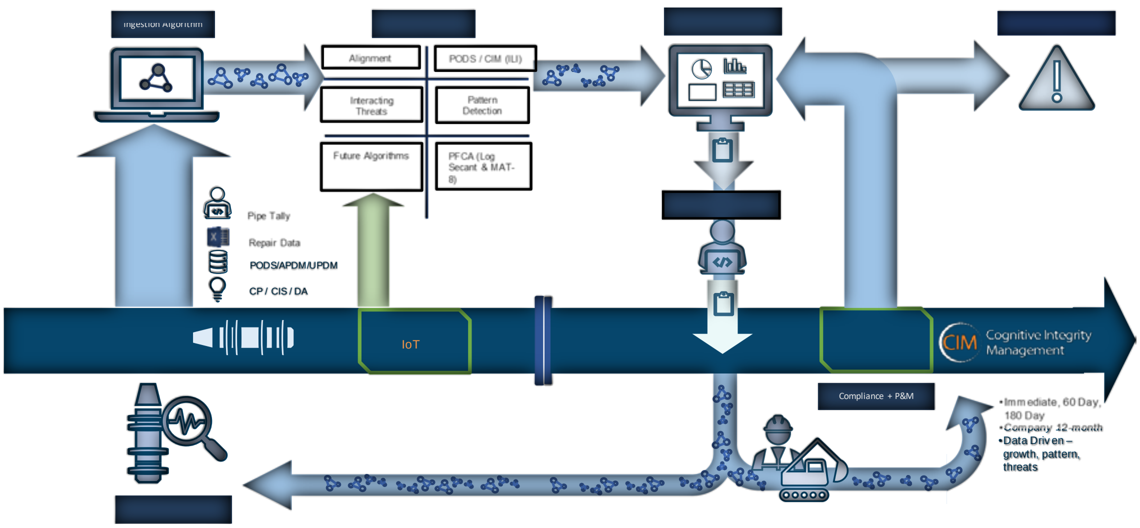


Figure 1 CIM Solution

Ingestion Algorithm

The data ingestion algorithms are based on a Bayesian classifier trained on data from over 5,000 ILI reports, and more than 50 million anomaly indications that have been gathered from customers utilizing Cognitive Integrity Management [3]. The solution interprets ILI vendor report formats and normalize these into structured datasets & schema. Pipeline operators have supported the mapping process of the columns on the ingested spreadsheets to CIM's Alias model, thus creating standardized "truth" data. For example, "depth (%)" from operator A and "depth (percent)" from operator B becomes "depth" for all operators. Eventually, all scenarios from each tool and vendor will be classified. The cloud enables the ability to do this without sharing confidential operator data. Figure 2 shows a snapshot of such ingestion analysis results with three In-Line Inspection (ILI) reports. For instance, the Above ground Markers were named as "AGM" in 2018 ILI and "Above Ground Marker" in 2008 ILI. Similarly, the Girth welds were named differently in different ILI reports, but all have been recategorized as Girth Weld in CIM.

The ingestion process also extracts the semantic meaning of vendor anomaly type and comment information into a standardized alias taxonomy, classification, category, and type structure. Again, data science is critical here. It can observe patterns in the data, for example, to extract the word "dent" from the comment field, tag the record, and update its alias to reflect this while maintaining the original user classification. The identified feature can now be used in Corrosion within Dent Interacting Threat algorithm.

Alias Type	1998ILI	2008ILI	2018ILI - 2018ILIWendy00000
AGM		**79**	**85**
Above Ground Marker		79	
AGM			85
Bend		**2,026**	**1,732**
BEND			1,732
Bend Begin		1,013	
Bend End		1,013	
Casing End		**10**	**12**
Casing begin		5	
Casing end		5	6
Casing Start			6
Clamp On Sleeve		**49**	**1**
Clamp		45	
Clamp begin		2	
Clamp end		2	
OTHE			1
Corrosion Wall Loss	**1,382**	**106**	**7,552**
ANOM	1,301		
Corrosion		94	7,217
Corrosion cluster		12	334
MIDE	1		
Total	**16,370**	**17,778**	**40,227**

Alias Type	2008 ILI	2018 ILI - 2018ILI60KSMYS	98ILI
AGM	**79**	**85**	
Above Ground Marker	79		
AGM		85	
Bend	**2,026**	**1,732**	
BEND		1,732	
Bend Begin	1,013		
Bend End	1,013		
Girth Weld	**14,986**	**15,075**	**14,988**
WELD	14,774	14,918	14,988
WELD-Change in Wall thickness	162	157	
Weld-Installation begin	13		
Weld-Installation end	13		
Weld-Iso joint begin	3		
Weld-Iso joint end	3		
Weld-Launcher end	1		
Weld-Other	16		
Weld-Receiver begin	1		

Figure 2 Creating "Truth" data using alias model

Automated girth weld alignment and anomaly alignment identifies geometric patterns in the data which can then be used to infer the most accurate matching of like features across independent datasets, including flow reversal where the values reported in one dataset are inverse with respect to another. The result is a spatially normalized representation of the data which supports growth analysis and integration of data from otherwise independent systems.Figure 3 shows the analysis results for a pipeline where the flow has been reversed in 2020 inspection compared to the 2010 inspection, yet the data was fully aligned by using the algorithm.

Assessment Name	ILI2010			ILI2020		
Master Joint ID	Joint No.	Log Distance	Joint Length	Joint No.	Log Distance	Joint Length
500,005,700.00	68070	1,972.39	58.56	68070	389,849.05	58.62
500,005,800.00	68060	2,030.95	60.16	68060	389,790.43	58.61
500,005,900.00	68050	2,091.11	57.96	68050	389,730.26	60.17
500,006,000.00	68040	2,149.07	59.73	68040	389,672.31	57.96
500,006,100.00	68030	2,208.80	46.11	68030	389,612.91	59.39
500,006,200.00	68020	2,254.92	59.42	68020	389,566.41	46.50
500,006,300.00	68010	2,314.34	57.44	68010	389,506.86	59.55
500,006,400.00	68000	2,371.78	59.53	68000	389,449.28	57.58
500,006,500.00	67990	2,431.31	59.48	67990	389,389.71	59.57
500,006,600.00	67980	2,490.79	58.55	67980	389,330.15	59.56
500,006,700.00	67970	2,549.34	57.66	67970	389,271.60	58.55
500,006,800.00	67960	2,607.00	57.69	67960	389,213.86	57.74
500,006,900.00	67950	2,664.69	57.50	67950	389,156.12	57.74

Figure 3 weld alignment for a reversed pigging scenario

Cognitive Learning

CIM leverages machine learning and applies an approach based on data science to the challenges of ingesting and normalizing a wide variety of integrity datasets such as ILI, GIS and asset data, CP survey, NDE, and repair data into a standardized structure which can then be aligned and analyzed. Algorithms based on pattern recognition are applied to these datasets first to identify and spatially align the mutually visible features present in each dataset such as block valve locations, girth welds, down to the individual geometric patterns of corrosion and ILI features. This enables the solution to identify and analyze how specific features are changing over time and supports the development of a comprehensive corrosion growth model. Finally, a suite of additional algorithms trained to identify specific patterns in corrosion geometry can help identify conditions such as areas of potential coating disbondment, internal localized corrosion, and seam weld corrosion.

Multiple corrosion growth models are supported based on a spatially normalized dataset including automated alignment of every pipeline anomaly reported throughout its complete assessment history.

A pit-to-pit depth comparison based on measured wall thickness over time can be used to determine a unique corrosion growth rate for each active anomaly in the pipeline system. This is one of the most accurate and least conservative method for calculating corrosion rates. An example is shown in Figure 4, where a single anomaly' growth has been detected in three consecutive ILIs.

	2018 ILI				2008 ILI				1997 ILI			
Corrosion Growth Rate (MPY)	Odometer (ft)	Length (in)	Width (in)	Depth (%)	Odometer (ft)	Length (in)	Width (in)	Depth (%)	Odometer (ft)	Length (in)	Width (in)	Depth (%)
-2.1	29,949.17	0.51	1.26	48.0%	29,958.41	0.43	1.93	57.0%	29,922.59	0.43	0.94	47.0%
6.2	257,024.30	4.13	2.91	40.0%	257,104.07	0.87	0.59	13.0%	257,291.05	0.32	0.39	8.0%
6.2	257,024.52	0.83	0.28	40.0%	257,104.07	0.87	0.59	13.0%	257,291.05	0.32	0.39	8.0%
6.8	265,221.24	3.90	4.13	44.0%	265,304.75	1.97	2.87	14.0%	265,501.16	0.32	0.32	8.0%
6.8	265,221.52	0.55	0.20	44.0%	265,304.75	1.97	2.87	14.0%	265,501.16	0.32	0.32	8.0%
4.6	265,280.65	9.06	14.80	40.0%	265,364.37	0.75	1.06	20.0%	265,560.71	0.43	0.39	7.0%
5.3	265,280.65	9.06	14.80	40.0%	265,364.38	0.79	0.83	17.0%	265,560.71	0.43	0.39	7.0%
6.8	265,280.65	9.06	14.80	40.0%	265,364.52	0.83	0.75	10.0%	265,560.71	0.43	0.39	7.0%
6.8	266,116.42	12.24	43.98	40.0%	266,200.80	1.14	0.55	10.0%	266,397.68	1.30	1.10	5.0%
7.1	266,116.42	12.24	43.98	40.0%	266,200.81	0.94	0.63	9.0%	266,397.68	1.30	1.10	5.0%
6.6	266,116.42	12.24	43.98	40.0%	266,200.82	1.65	0.87	11.0%	266,397.72	0.83	1.30	6.0%
7.5	266,116.42	12.24	43.98	40.0%	266,200.82	0.71	1.18	7.0%	266,397.72	0.83	1.30	6.0%
6.2	266,116.42	12.24	43.98	40.0%	266,200.82	0.79	0.55	13.0%	266,397.72	0.83	1.30	6.0%
7.1	266,116.42	12.24	43.98	40.0%	266,200.83	0.71	0.98	9.0%	266,397.72	0.35	0.39	8.0%
8.0	266,116.42	12.24	43.98	40.0%	266,200.83	0.83	5.51	5.0%	266,397.72	0.35	0.39	8.0%
5.9	266,116.42	12.24	43.98	40.0%	266,200.83	0.83	0.55	14.0%	266,397.75	0.55	0.55	5.0%
7.3	266,116.42	12.24	43.98	40.0%	266,200.92	0.51	0.55	8.0%	266,397.75	0.55	0.55	5.0%
7.5	305,024.68	3.94	3.39	42.0%	305,138.00	1.22	13.78	9.0%	305,448.56	0.20	0.20	8.0%
7.5	305,024.99	0.28	0.91	42.0%	305,138.00	1.22	13.78	9.0%	305,448.56	0.20	0.20	8.0%

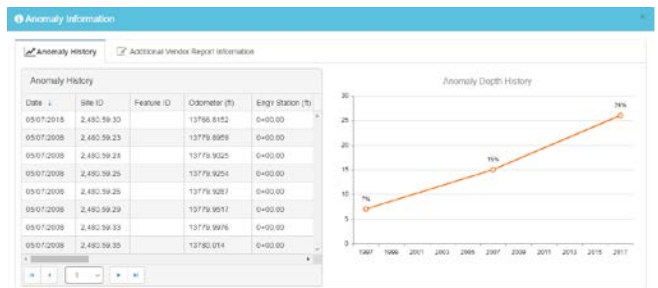


Figure 4 Pit to pit corrosion growth calculation using 100% of data set

Pit to pit corrosion growth calculation works by directly comparing measured wall thickness changes after a known time interval. In locations where the pit-to-pit growth is not present, the solution falls back to half life calculation which is the second most accurate method to calculate corrosion growth where future inspections are managed based on the worst-case half life established at each location.

Business Intelligence

The insights resulting from cognitive learning and algorithmic processing of pipeline integrity data can then be combined into a comprehensive analytical data model which supports business intelligence queries and advanced analysis capabilities. By combining data from ILI, CP Survey, GIS, and dig site into a single source, improved forecasting capabilities, predictive maintenance, and data-driven decision making across multiple levels of the organization can be enabled with a robust business intelligence platform. Figure 5 shown an example of a pipeline with three ILIs where a high-level overview of the integrity data is provided.

Data which has been integrated and correlated from a variety of sources yield advanced insights which would be otherwise invisible when looking at each of the datasets independently. Changes in patterns, trends over time, and meaningful insights can be surfaced readily by interactive reporting and data visualization tools, including the ability to understand asset integrity across the entire organization and to drill down into an individual system and individual pipeline anomalies in intricate detail. Through rich 3D visuals, geometric patterns and interactions can be easily observed including corrosion along a spiral seam weld, corrosion affecting a girth weld or HAZ, or concentrated in an area or orientation on the pipeline.

Figure 5 Interpreting and visualizing integrity data using Power BI

Integrity Threat Identification and Pattern Detection

Through implementation of business intelligence, the integrity engineer will be able to have a detailed review of each anomaly and joint with respect to the location and history of anomalies. Two examples are provided in Figure 6. The pipeline in left side has a pattern of corrosion at the bottom of the pipe and the right side of the figure shows the detail at anomaly level where the two consequent ILIs had identified two different set of anomalies using two different tool technologies, i.e., MFL A and MFL C.

Figure 6 Visualization of anomalies and pipe tally at anomaly/pipe joint level

The ability to visualise the data and align to the past inspection enables the integrity engineers to identify inconsistency with the inspection reports. Figure 7 shows an example of a pipeline with 3 ILIs reports. The 2016 and 2018 ILIs consistently show a pattern of bottom of the line corrosion where the 2020 report shows a change of pattern from bottom of the line to top of the line. This allowed the integrity engineer to assess the quality of the ILI data based on historical evidence.

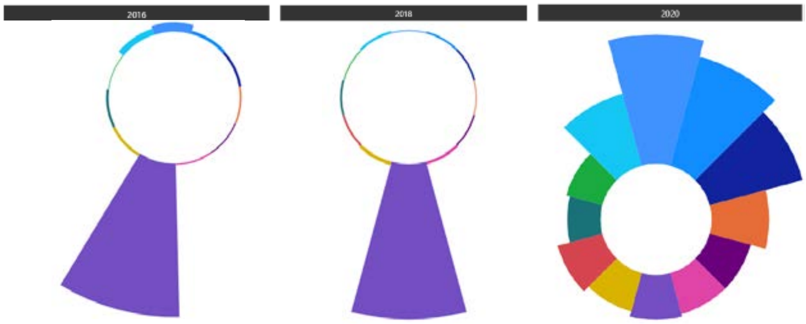


Figure 7 Inconsistency with Anomaly location in latest ILI vs historic ILIs

ML correlates the data enabling the integrity engineer to efficiently interrogate the inspection data and identify future threats. Figure 8 shows an example of a newly built pipeline where on number of joints, an identical pattern was found. Correlating with the pipe materials where a spiral welded pipe was used, the integrity engineer was able to identify future integrity threat.

Figure 8 A pattern of anomalies across spiral weld in a newly built pipeline & A pattern of anomalies on field joint on a 30 year old pipeline

Figure 8 also shows another example of a 30 years old pipeline where ML enable the integrity engineer to identify the main integrity threat to the pipeline field joint coating failure.

These examples demonstrate how ML can be utilized to focus its future inspection and mitigation strategies at the right place and in the right time.

Fitness for Service Assessment using ML

Integrity engineers are challenged with performing an accurate and timely, but also cost-effective, fitness for service assessment upon receiving inspection results and developing a forward maintenance program. This is where ML's capability in leveraging on its speed, using aligned pit to pit growth calculations and pre-defined fitness for service assessment criteria, can play a significant role in saving on cost and time. The analysis can be set to use certain criteria or conditions as shown in Figure 9. The integrity engineer would be able to run as many hypotheses as necessary to identify the most credible scenarios to develop an integrity management program. Figure 9 shows an example where a scenario was considered for a gas pipeline with no ILI tool tolerance using a pit to pit corrosion growth calculation under certain conditions as shown.

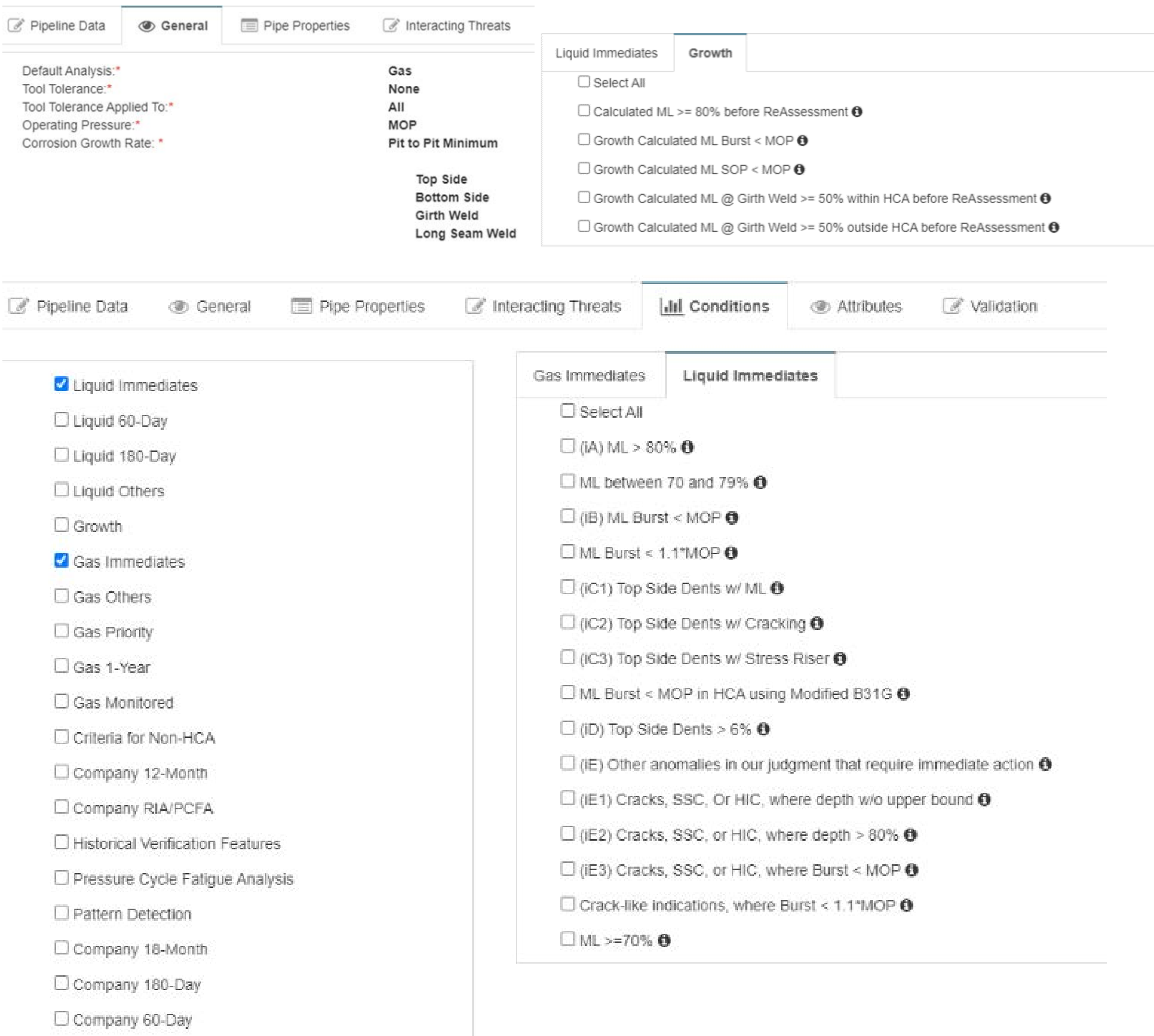


Figure 9 Analysis setup

CIM has more than 200 defined FFS assessment conditions built into the solution covering regulatory requirements, standard requirements such as ASME B31.G and company requirements.

Upon completing the analysis, the locations/anomalies/joints where an integrity action is required to be undertaken exceeding certain criteria will be identified as shown in Figure 9.

The integrity engineer would be able to review all information related to each individual anomaly in one screenshot and make the decision on the action required.

Integrity Management and Dig Up

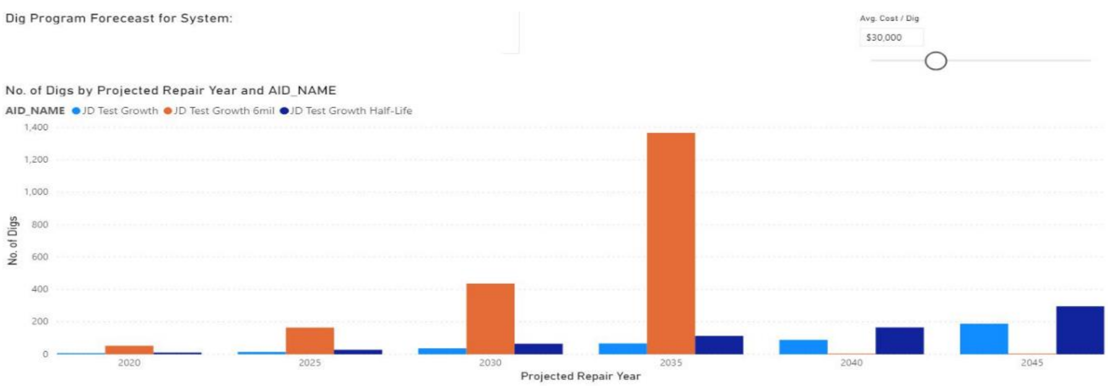
Upon completion of all assessment the integrity engineer will be able to develop a short term and long-term integrity plan for pipeline.

Using machine learning and the ability to correlate the integrity data enable the integrity engineers to maximize the benefits and significantly reduce the integrity activities and dig up costs. Figure 10 shows an example where the application of ML had significantly reduced the cost of maintenance for a 30 year old pipeline. The estimated cost of maintenance in 5 years is reduced by a factor of 4. What the operator predicted as cost in the next 5 years is 75% higher than what ML predicted for the next 10 years. ML also reduced the ILI frequency as less digs are required using ML.

		3 years		5 years		10 Years	
		CIM	Existing Calculation	CIM	Existing Calculation	CIM	Existing Calculation
# of Digs		**4**	**5**	**6**	**26**	**12**	**3482**
Threat type	1	4	3	5	6	6	207
	2	0	1	0	4	2	551
	3	0	1	1	16	4	2724

Figure 10 Case Study - Comparing ML results with traditional methods

The predominant reason for reducing the cost of integrity dig up using ML is the pit-to-pit growth measurement capabilities. Figure 11 shows the difference in a pipeline in one of the case studies. The cost of dig up program exponentially increases with a flat rate calculation even if the method is segment based.



AID_NAME	JD Test Growth		JD Test Growth 6mil		JD Test Growth Half-Life	
Projected Repair Year	No. of Digs	Dig Program Cost	No. of Digs	Dig Program Cost	No. of Digs	Dig Program Cost
2020	7	$210,000	53	$1,590,000	11	$330,000
2025	15	$450,000	165	$4,950,000	28	$840,000
2030	37	$1,110,000	436	$13,080,000	65	$1,950,000
2035	67	$2,010,000	1366	$40,980,000	114	$3,420,000
2040	89	$2,670,000	1	$30,000	166	$4,980,000
2045	189	$5,670,000	1	$30,000	296	$8,880,000
Total	**404**	**$12,120,000**	**2022**	**$60,660,000**	**680**	**$20,400,000**

Figure 11 comparing pit to pit growth measurement with half-life growth and a flat rate calculation.

Regulatory Compliance

There are more than 200 conditions in the library for liquid and gas integrity concerns. The library includes CFR 192 & 195 regulations, industry best practices, CIM data science and machine learning patterns, and company specific conditions.

Regulatory compliance can represent a significant expense and resource commitment for pipeline operators. Through capturing the assessment planning activity, integrity decision-making, and mitigation resulting from those decisions in a comprehensive integrity data system it becomes trivial to produce annual regulatory reporting and to comply with audit requirements. Correlated data sets result in savings for regulatory reporting functions, such as CIM's automated generation of annual F and G reports required by PHMSA, at a press of a button rather than by manual, onerous processes. Layering of seemingly disparate datasets also allow CIM to automate detection of >200 threat detection patterns, which is impossible to find with human manual processes

Integration of Other Inspections

The core algorithm in CIM is alignment algorithm. More than 22 releases of this algorithm have resulted in some very significant learnings that have facilitated alignment at the system level – that is, all ILI assessments ever done on a system. The algorithm was leveraged as the baseline for the ILI-to-GIS (PODS) alignment within CIM, which has been extended to include mutually visible assets (valves, casings, AGMs, etc.) that provide spatial (latitude, longitude, elevation)-to-engineering stationing (begin/end, m-value) alignment.

As the data captured from CP surveys is limited in depth (on, off, dBµV, %IR, depth, notes, latitude, longitude, elevation, and a few others), there are only a few options to align it to existing ILI and/or GIS (PODS) data. It utilizes the spatial latitude and longitude coordinates to snap it to the nearest ILI location. As CIS measurements are typically taken every 5-10 feet and ACVG/DCVG indications occur as coating anomalies are identified, this process will overlay multiple CIS readings and possibly multiple ACVG/DCVG indications onto a single joint of pipe [5]. CIM then correlates each anomaly (corrosion, dent, crack) that has already had pit-to-pit alignment across all ILI assessments done to the closest CP survey results point. It can then overlay rate of corrosion growth and anomalous CP patterns.

CIM now supports and manages data from ILI, GIS (PODS), Repair, and CP surveys that have been aligned at the anomaly level where there is sufficient data across an operator's entire pipeline as shown in Figure 12

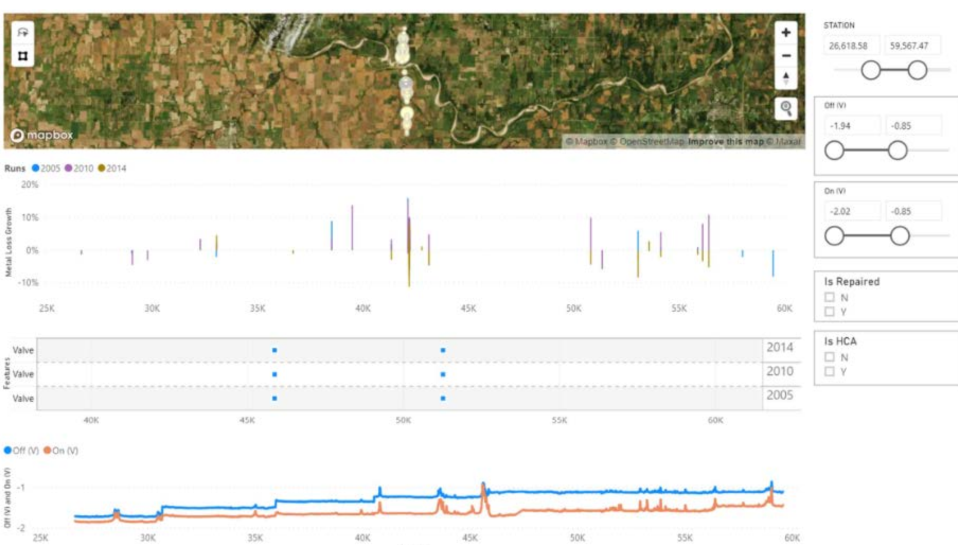


Figure 12 Integration of different inspection results

Conclusion

This paper presented how machine learning can be used in pipeline integrity engineering using a newly developed ML based tool. It has explored methodologies around how business intelligence is derived from machine learning and data science analysis. Using ML, this paper demonstrates how an integrity engineer would be able to:

- Review all the anomaly and feature data on entire pipeline system.
- Gather insights from interacting threats and pattern detection, rather than relying on single data points.
- Gain visibility into the entire pipeline—bubble up and drill down into the finest level of detail.

Leverage on 100% of the existing data set using ML enables the integrity engineer to use the most advanced corrosion growth rate calculation, i.e., a pit to pit corrosion growth rate. The case study found that the estimated cost of maintenance in 5 years was reduced by a factor of 4 and what the operator predicted as cost in the next 5 years was 75% higher than what ML predicted for the next 10 years. This has significantly contributed in managing pipeline integrity more effectively and with a significantly reduced cost of integrity dig up and consequently changing the pipeline integrity strategy from reactive to proactive.

References

[1] M. P. P. V. P. Gerhardus H. Koch, "Corrosion Costs and Preventive Strategies in the United States', Report FHWA-RD-01-156, NACE International,".

[2] "New Direction, Complex Choices, The outlook for the il and gas industry in 2020, DNV GL,OSLO,2020".

[3] J. D. M. H. M. T. E. Yevgeniy Petrov, "STATISTICAL ANALYSIS OF DIG OPERATIONS LEADING TO PRODUCTIVE REPAIRS," in *Proceedings of the 2020 13th International Pipeline Conference*, Calgary, 2020.

[4] J. D. M. H. M. T. E. Yevgeniy Petrov, "Statistical analysis of dig operations leading to productive repairs," in *Pipeline Pigging and Integrity Management Conference*, Houston, 2020.

[5] OneBridge Solutions Inc., "Data Integration - Cathodic Protection Survey," OneBridge Solutions Inc., Boise, 2020.

