DOT-PHMSA Contract # 693JK31910015POTA

Improving the Reliability, Detection and Accuracy Capabilities of Existing Leak Detection Systems (CPMS) Using Machine Learning

PRCI Project PL-1-07

David Vickers, Southwest Research Institute

Heath Spidle, Southwest Research Institute

March 17, 2022

LEADING PIPELINE RESEARCH

Project Team

- David Vickers – SwRI
- Heath Spidle – SwRI
- Carrie Greaney – PRCI
- Chevron
- TC Energy
- Enterprise
- Enbridge
- Plains All-American Pipeline

Challenges and Goals Addressed

This project looked to address three primary leak detection systems gaps: ability to (1) find smaller leaks, (2) find leaks faster and (3) find leaks more reliably (higher confidence, lower false alarms) than is possible today with conventional CPM systems.

Goals:

- **Increase the sensitivity of existing CPM leak detection systems to smaller leaks**
- **Decrease the reaction time of the system to anomalous events both onset and resolution**
- **Reduce false positive alarms due to transient events**

Tasks & Timeline

Technical and Deliverable Milestone Schedule					
Item No.	**Task No.**	**Activity/Deliverable**	**Expected Completion Date/Mos**	**Payable Milestone**	***Total**
	(per proposal)	ACTIVITY/DELIVERABLE		TITLE	
1	1	Operator Survey Documentation	1 months	Operator Survey Documented in 1st Quaterly report	17,017.00
2	5	1st Quarterly Status Report	3 months	Submit 1st quarterly report	7,650.00
		First Payable Milestone	3 months	SUBTOTAL	24,667.00
3	2	Data Curation and Analysis Documentation	3 months	Data Curation and Analysis Completion - Summary provided in 2nd Quaterly Report	61,816.00
4	5	2nd Quarterly Status Report	3 months	Submit 2nd quarterly report	7,650.00
		Second Payable Milestone	6 months	SUBTOTAL	69,466.00
5	5	3rd Quarterly Status Report	9 months	Submit 3rd quarterly report	85,565.00
		Third Payable Milestone	9 months	SUBTOTAL	85,565.00
6	3	ML Agorithm Development Summary Documentation	7 months	ML Algorithm Development Documented in 4th Quaterly Report	85,697.00
7	5	4th Quarterly Status Report	12 months	Submit 4th quarterly report	7,650.00
		Fourth Payable Milestone	12 months	SUBTOTAL	93,347.00
8	5	5th Quarterly Status Report	15 months	Submit 5th quarterly report	74,739.00
		Fifth Payable Milestone	15 months	SUBTOTAL	74,739.00
N/A	N/A	6th Quarterly Status Report	18 months	Submit 6th quarterly report	0.00
		6th Payable Milestone	15 months	SUBTOTAL	0.00
N/A	N/A	7th Quarterly Status Report	21 months	Submit 7th quarterly report	0.00
9	4	ML Algorithm Testing and Validation - AI Framework and Guidance Document	22 months	AI Framework and Guidance Document Delivery	0.00
10	5	Prepare and Submit Draft Final Report	22 months	Submit draft final report	7,650.00
				GRAND TOTALS	347,784.00

AVAILABLE DATA

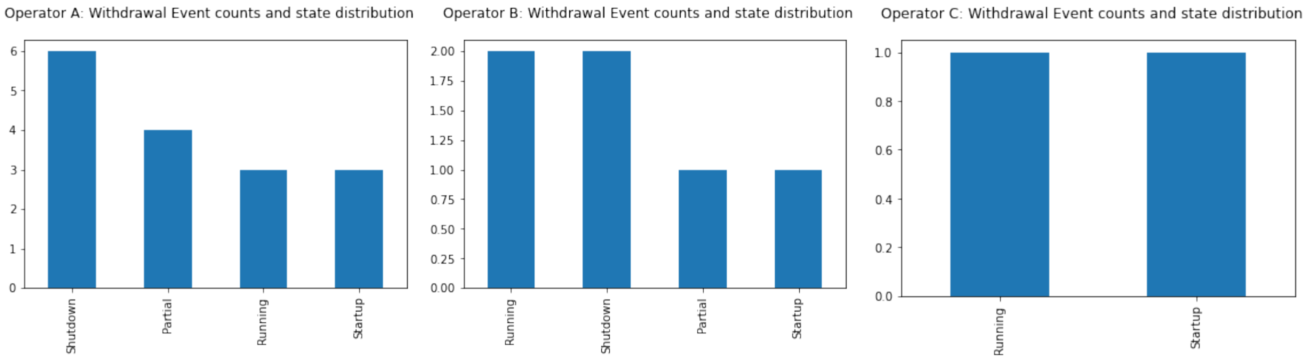
Data Summary

Vendor	Length of Data	Number of Data Samples	Number of Commodity Withdrawal Tests	Number of Pipeline Segments	Withdrawal Rates	Withdrawal Rate as Percent of Flowrate *
Operator A	3 Days	3,181	16	5	16.5 bph (2.6 m3h) to 150 bph (24 m3h)	0.5% to 4.5%
Operator B	365 Days	524,106	2	2	20 bph (3.2 m3h) to 100 bph (16 m3h)	0.4% to 2.5%
Operator C	10 Days (non-contiguous) 3 March (withdrawal), 7 Aug	172,790	6	9	55 m3h (340 bph) to 130 m3h (812.5 bph)	1.2% to 3%
Totals	1 year 13 days	700,077	24	16		

*** Percentage flowrates are based on commodity withdrawal rate and flow rate at the time of the withdrawal, not on pipeline maximum or designed specifications.**

Pipeline State and Withdrawal Representation

- **Representative withdrawal events constitute 0.3% of the total data received.**
- **Each operator's dataset lends itself better to a different type of architecture**
  - Operator A: Several Examples of withdrawals during different operational states
  - Operator B, C Lots of Operational data with a few withdrawals



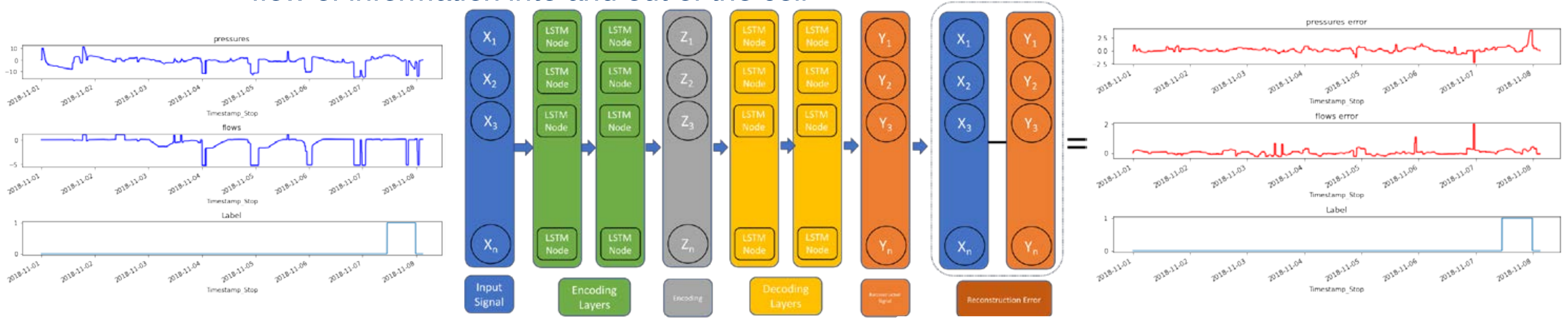
MACHINE LEARNING

Algorithm Development

- **Explored several machine learning and deep learning techniques**
  - Un-Supervised techniques due to large class imbalance
  - Supervised techniques with data augmentation and down sampling
- **Explored decoupling the relationship between operational states of the pipeline and representative withdrawals is a challenge**
  - Limited information on withdrawals during various operational states
- **Explored Ways of segmenting data so one model can be trained on all data for generalizable deployments**
- **Explored Standardization and Normalization and data processing of each set and how it relates to the effects of pipeline states**

Un-Supervised Learning

- **Multivariate Time Series LSTM Autoencoder**
  - Unsupervised anomaly detection is useful when there is little to no information about anomalies and related patterns
  - Long-Short Term Memory (LSTM)
    - *A common LSTM unit is composed of a **cell**, an **input gate**, an **output gate** and a **forget gate**. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell*

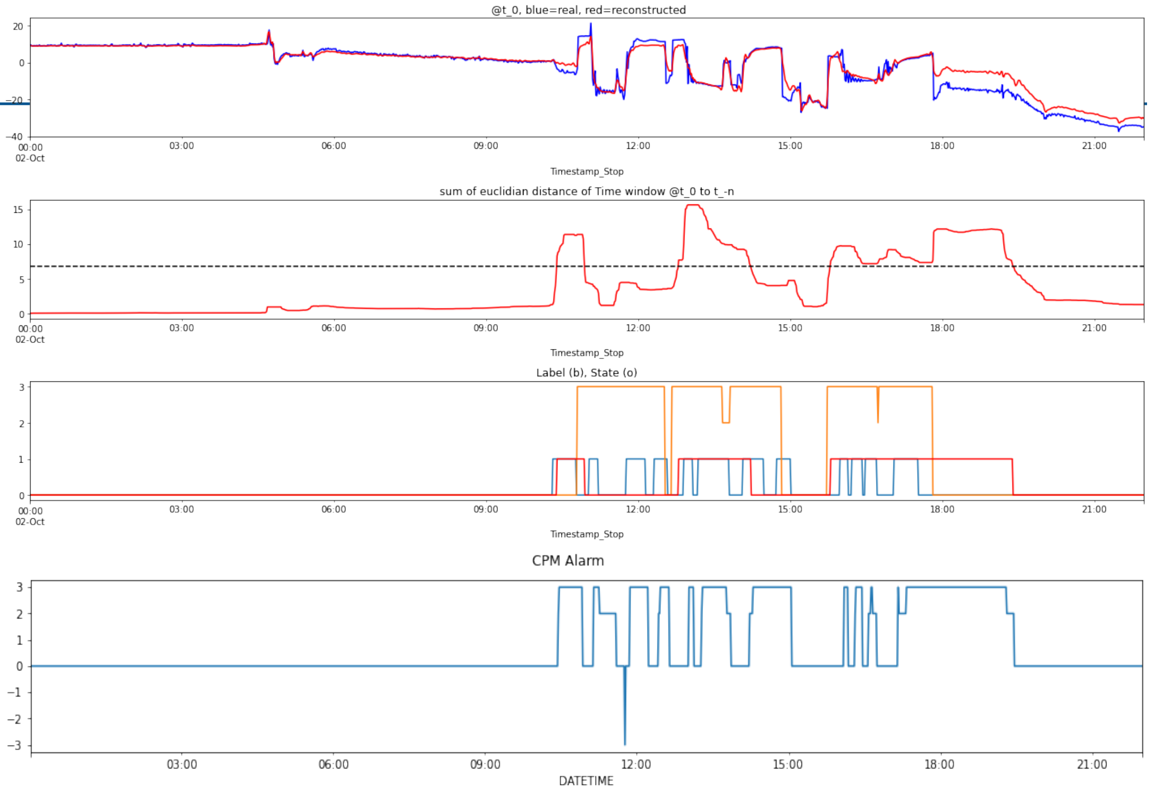


RESULTS

Operator A

- **CPM Accuracy: 82%**
- **ML Accuracy: 78%**

	CPM System	Autoencoder
True Negative	1016	883
False Positive	188	203
False Negative	77	92
True Positive	157	142
Weighted F1	0.83	0.79
Accuracy	0.82	0.78



Operator B

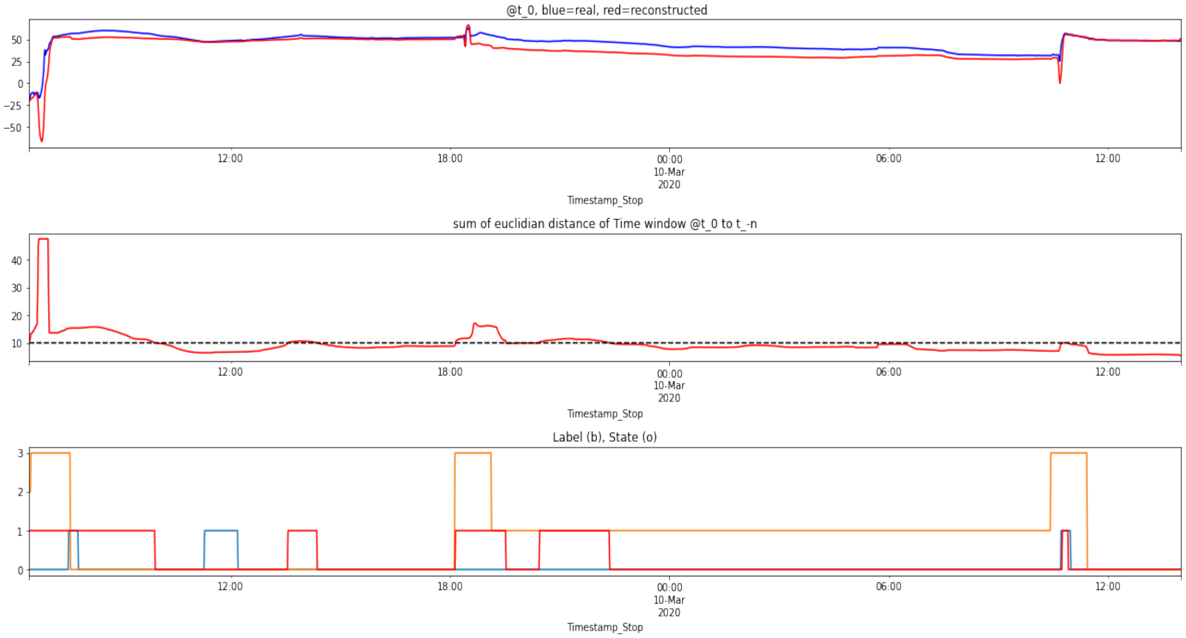
- **CPM Accuracy: 43%**
- **ML Accuracy: 73%**

	CPM System	Autoencoder
True Negative	658	852
False Positive	193	0
False Negative	730	443
True Positive	36	323
Weighted F1	0.34	0.70
Accuracy	0.43	0.73

Operator C

- **No CPM data**
- **ML Accuracy: 74%**

	Autoencoder
True Negative	1532
False Positive	487
False Negative	10
True Positive	11
Weighted F1	0.85
Accuracy	0.76



Comparison

	Operator A		Operator B		Operator C
	CPM System	Autoencoder	CPM System	Autoencoder	Autoencoder
True Negative	733	542	1674	1867	1532
False Positive	39	111	704	485	487
False Negative	53	62	0	49	10
True Positive	75	66	89	40	11
Weighted F1	0.90	0.79	0.80	0.85	0.85
Accuracy	0.90	0.78	0.71	0.78	0.76

FRAMEWORK

Machine Learning Framework and Guidance

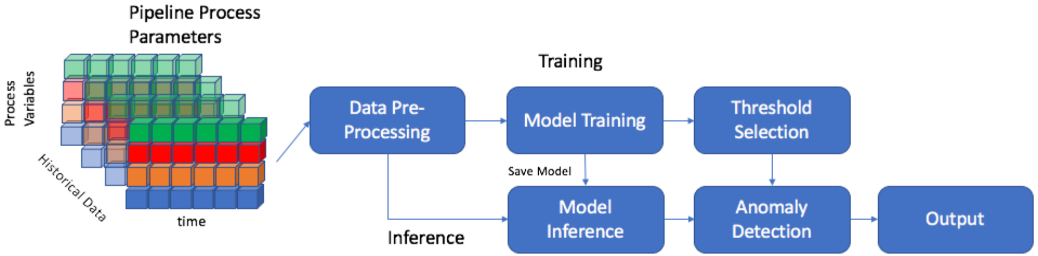
- **An output of this effort is a framework and code base that can be used by operators to train a model on their current pipeline setups and compare the performance of the anomaly detection network with their CPM system**
- **”Machine Learning Guidance and Framework” Document provided to members**
  - How to format data
  - How to train a model
  - How to evaluate a model
  - How to run in real-time

Training

- **Data**
  - Historical pipeline data in tabular format
  - Pipeline Metadata
    - *Pressure Readings*
    - *Flow Rate Measurements*
    - *Segment Inputs/Outputs*
- **Results:**
  - Trained Anomaly Detection Model on section of pipeline

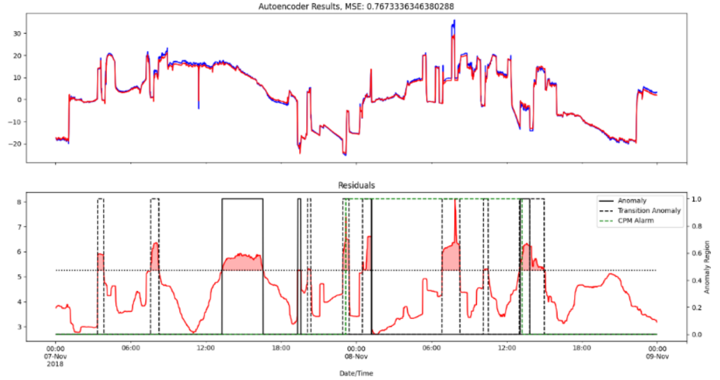
DateTime	Pressure_Out-Station-1	Pressure_In-Station-2	Pressure_Out-Station-2	Pressure_In-Station-3	FlowRate-Station-1	FlowRate-Station-4	FlowRate-Station-2	FlowRate-Station-5	Station-1-Pump1	Station-1-Pump2
2017-10-01 10:00:00	985	532	1037	440	3602	555.03	1286	665	2	2
2017-10-01 10:01:00	985	532	1037	440	3597	554.87	1283	665	2	2
2017-10-01 10:02:00	985	532	1037	440	3599	554.55	1286	665	2	2
2017-10-01 10:03:00	985	532	1035	440	3602	554.55	1283	665	2	2
2017-10-01 10:04:00	985	532	1035	440	3602	554.55	1288	665	2	2

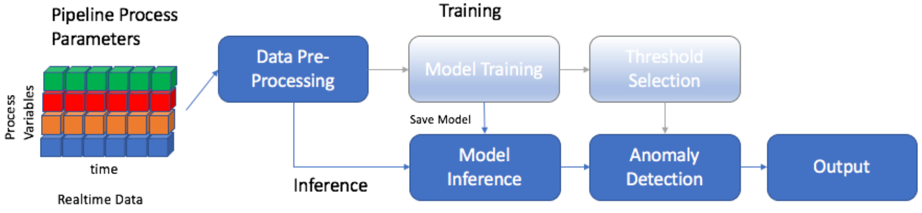

```
{
    "Pressure_Sections" : [
        [
            "Pressure_Out-Station-1",
            "Pressure_In-Station-2"
        ],
        [
            "Pressure_Out-Station-2",
            "Pressure_In-Station-3"
        ]
    ],
    "Input_FlowRates" :[
        "FlowRate-Station-1",
        "FlowRate-Station-4"
    ],
    "Output_FlowRates" : [
        "FlowRate-Station-2",
        "FlowRate-Station-5"
    ],
    "Pump_Status" : [
        "Station-1-Pump1",
        "Station-1-Pump2"
    ]
}
```

Validating Trained Model

- Data:
  - Historical pipeline data in tabular format
  - Same metadata as training data
  - Optional:
    - CPM System alarm data for historical period
- Output:
  - Visual Plots showing anomalous regions in the data and corresponding ground truth





Realtime Anomaly Detection

- **Inputs:**
  - Trained pipeline anomaly model
  - Data:
    - *Live MQTT stream*
    - *Tabular data*

- **Realtime display of prediction errors, and alarming threshold**

Conclusion

- The presented research and ML implementations have, in some cases, demonstrated slightly more sensitivity to withdrawals than the current CPM system integrations and, in most cases, at least as sensitive to withdrawal as the current CPM systems for the two pipelines in which corresponding CPM outputs were provided.
- It is anticipated that operators will be able to use the AI framework developed in this project in accordance with the guidelines provided to generate models which can be integrated with their leak detection systems to increase their sensitivity and decrease false alarms.

Next Steps

- **Three Follow-on proposals submitted to PRCI**
  - Liquid Pipeline CPM System Machine Learning Phase 2: Thresholding Improvement
    - *incorporate and investigate further datasets that survey a pipeline over long periods of time, so that the algorithms have the ability to model nominal pipeline behavior in a variety of conditions.*
  - External Leak Detection
    - *The goal of the proposed research is to use visible and thermal cameras to autonomously detect liquid crude oil leaks both pooling and spraying.*
  - Natural Gas Pipeline Monitoring using Machine Learning
    - *The goal of the proposed research is to transfer the ideas and lessons learned from working with liquid pipelines and apply the techniques to gas pipeline measurement errors and associated Lost and Unaccounted For (LUAF) data.*

QUESTIONS?

Final Report Available on DOT-PHMSA Website:

https://primis.phmsa.dot.gov/matrix/PrjHome.rdm?prj=859

David Vickers, P.E.
Institute Engineer
David.vickers@swri.org

Heath Spidle
Research Engineer
Heath.Spidle@swri.org

Carrie Greaney
Sr. Program Manager
cgreaney@prci.org