**CAAP Quarterly Report**

**7/12/2023**

*Project Name: Pipeline Risk Management Using Artificial Intelligence-Enabled Modeling and Decision Making*

*Contract Number: 693JK32150001CAAP*

*Prime University: Rutgers University*

*Prepared by: Bingyan Cui (PhD student), Emad Farahani (PhD student), Dr. Hao Wang (PI), Dr. Qindan Huang (Co-PI)*

*Reporting Period: 4/1/2023 – 6/30/2023*

**Project Activities for Reporting Period:**

*Task 1 Literature Review (Completed)*

*Task 2 Data Collection from Industry Partners (Completed)*

*Task 3 Data-Driven Probabilistic Modeling of Pipeline Defects*

***Predicting Defect Growth Using Bayesian Neural Network***

To improve the prediction performance of BNN, the tuning process includes selecting appropriate hyper-parameters and adjusting the network architecture were conducted. The hyper-parameters were optimized, including learning rate and batch size. On the other hand, different model architectures were tried, such as changing the number of layers and the number of units in each layer. The obtained model parameters after tuning is shown in Table 1.

Table 1. Model parameters used in BNN after tuning

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Defect type | Learning rate | Batch size | No. of layers | No. of units in each layer |
| Metal Loss | 0.001 | 69 | 2 | (64, 16) |
| Defect Length | 0.001 | 52 | 2 | (32, 8) |

The results show that the fitting results of BNN model are improved through tuning process. The overall comparison of model accuracy in terms of R-square values before and after BNN tuning is presented in Table 2.

Table 2 Comparison of model accuracy before and after BNN tuning

|  |  |  |
| --- | --- | --- |
| Soil Property Input | Defect Type | BNN |
| Before | After |
| Original dataset | Metal Loss | 0.49 | 0.58 |
| Defect Length | 0.35 | 0.38 |
| Zone-based dataset | Metal Loss | 0.64 | 0.77 |
| Defect Length | 0.94 | 0.96 |

Fig. 1 shows the BNN model prediction results using the zone-based dataset.



(a) (b)



(c) (d)

Fig. 1 BNN model results after tuning using the zone-based dataset: (a) prediction of metal loss; (b) prediction of defect length; (c) predicted vs. measured for metal loss; (d) predicted vs. measured for defect length.

***Introducing Corrosion Time Parameter into Model***

In the above analysis, the pipeline age was used as an input variable to predict metal loss and defect length. However, not all corrosion defects start at the installation time of pipeline. Therefore, using age parameter is not the best choice for predicting corrosion growth. Considering the corrosion defect grows over time, the parameter of corrosion time is introduced to replace the age parameter. The corrosion time is defined as the difference between the pipeline age and the initiation time of defect. The initiation time of defect is an essential factor when predicting damage growth in pipelines because it represents the point at which the damage process starts.

However, the accurate initiation time of corrosion is not available from the original dataset. Therefore, the initiation time is estimated based on the assumption that a defect with larger corrosion size occurs earlier than the others with smaller corrosion sizes. This methodology assumes non-uniform initiation time for all defects. It considers the possibility that different defects may have different initiation times, providing more realistic representation of the damage growth process.

The detailed process to determine the initiation time is as follows:

1. Start by assuming that the occurrence of defects follows a homogeneous Poisson process. This process models the random occurrence of defects in the pipeline over time.
2. Then, the initiation time of each individual defect is assumed to follow a Gamma distribution. The Gamma distribution is characterized by two parameters: shape parameter (*α*) and scale parameter (*β*). The initiation time of each defect follows a Gamma distribution, where the scale parameter (*β*) is treated as an unknown model parameter to be estimated, and the shape parameter (*α*) is assumed to be the ranking of each defect. This means that the higher the ranking, the earlier the initiation time.
3. Estimate the rate parameter *λ* using historical data on the occurrence of defects in the pipeline. Therefore, the scale parameter *β* can be estimated based on rate parameter *λ.* Taking into account the assumption that defects with larger sizes occur earlier, the shape parameter *α* can be regarded as the same of metal loss distribution.
4. Calculate the initiation time of each defect using the estimated *α* and *β* values and the Gamma distribution formula. This will provide a probability distribution for the initiation time of each defect, considering their ranking based on detected dimensions.

Based on the metal loss data, the frequency distribution of generated initiation time for all corrosion defects is shown in Fig. 2.



Fig. 2 Frequency distribution histogram of initiation time (year).

After replacing the age parameter by corrosion time parameter, the updated BNN model performance can be seen in Fig. 3, for the zone-based dataset.



1. (b)

 

(c) (d)

Fig. 3 BNN fitting results using the zone-based dataset (a) prediction of metal loss; (b) prediction of defect length; (c) predicted vs. measured for metal loss; (d) predicted vs. measured for defect length.

The comparison of model accuracy before and after using corrosion time parameter in BNN models is presented in Table 3.

Table 3 Comparison of R2 before and after introducing corrosion time

|  |  |  |
| --- | --- | --- |
| Soil Property Input | Defect Type | BNN |
| Before | After |
| Original dataset | Metal Loss | 0.49 | 0.84 |
| Defect Length | 0.35 | 0.42 |
| Zone-based dataset | Metal Loss | 0.64 | 0.90 |
| Defect Length | 0.94 | 0.95 |

Therefore, the inclusion of corrosion time into the BNN model provides more accurate prediction of defect growth in pipelines, taking into account the non-uniformity of corrosion initiation time and the ranking of defects based on their dimensions.

***Identification of New Defect Database***

Based on analysis results using the ILI data from the industry partner, the general trend of defect growth can be captured with the average corrosion depth and length. However, relative large variations were found in the prediction model results. This is probably due to the fact that the ILI data are only available for two inspections. Therefore, the research team conducted extensive literature search to identify the multiple-year defect database in the literature.

The identified new database is published by Velázquez et al. (2010). This database includes 259 maximum pitting corrosion depth data together with soil properties (redox potential, pH, soil resistivity, soil textural class, water content, bulk density, dissolved chloride, bicarbonate, and sulfate ion concentrations), pipe-to-soil potential, and coating type/condition. The data were measured and collected over a three-year period at the excavation locations on different onshore buried pipelines operating in southern Mexico for up to 50 years.

It is noted that the new database covers more comprehensive information for prediction of corrosion growth, including: 1) multiple-year inspections; 2) different soil types and comprehensive soil properties; 2) CP potential; and 4) coating type and condition. The limitation is that the lengths of defects were not reported and the data points are relatively small.

*Velázquez, J.C., et al., Field Study—Pitting Corrosion of Underground Pipelines Related to Local Soil and Pipe Characteristics. Corrosion, 2010. 66(1): p. 016001-016001-5.*

**Project Activities with Cost Share Partners:**

Cost share is provided by Rutgers University and Marquette University during this quarterly period as budgeted in the proposal.

**Project Activities with External Partners:**

The industry partner reviewed the draft paper “Analysis and Prediction of Pipeline Corrosion Defects Based on Data Analytics of In-Line Inspection” and agreed to submit the paper, which has been published in *Journal of Infrastructure Preservation and Resilience*.

**Potential Project Risks:**

N/A

**Future Project Work:**

Considering the importance of accurately predicting defect growth trend for repair scheduling and the identification of new database with more comprehensive information, the research team will continue working on Task 3 Data-Driven Probabilistic Modeling of Defects. Both BNN models and probabilistic power-law models will be used to analyze corrosion growth using the new database to evaluate the applicability and accuracy of models. And the Task 4 Quantification of Probability of Failure will be started after next quarter.

**Potential Impacts to Pipeline Safety:**

The ILI data will be used to develop probabilistic growth models of pipeline corrosion defects, which can aid pipeline operators better predict failure risk and make repair decisions.