CAAP Quarterly Report (April-June 2020)

Date of Report:	July 22, 2020
Prepared for:	U.S. DOT Pipeline and Hazardous Materials Safety Administration
Contract Number:	Cooperative Agreement #693JK31850011CAAP
Project Title:	Development of a prediction model for pipeline failure probability based on learning from past incidents and pipeline specific data using artificial neural network (ANN)
Prepared by:	Texas A&M Engineering Experiment Station
Contact Information:	Joshua Arnold
For quarterly period ending:	June 30, 2020

Business and Activity Section

(a) Contract Activity

The contract activities have been continued.

(b) Status Update of Past Quarter Activities

Dr. Noor Quddus (Assistant Research Engineer) has been leading the team for the project. Graduate students working in the project are Mengxi Yu (4th year PhD candidate) and Pallavi Kumari (3rd year Ph.D. student). An undergraduate researcher, Mason Boyd, has been working since last Summer. Syeda Zohra Halim, a postdoctoral research associate, also worked in this quarter for the project.

(c) Cost share activity

Due to some unforeseen difficulties, financial information is not available at this moment. In the next quarter, all cost sharing activity will be updated.

(d) Task 1: *Development of methodology for creating root cause analysis reports*
Task 2: Selection of training samples and development of the learning algorithm

Both Task-1 and Task-2 have been continued.
Detailed discussion and descriptions for the following:

1. Background and Objectives

Pipeline incidents data obtained from the PHMSA incident database and incident investigation reports are being analyzed to develop a predictive model based on artificial neural network. Since incident reports collect only one apparent cause or root cause and one sub-cause per incident, the interdependency or relationship among the causal factors cannot be identified. Incident descriptions and incident investigation reports are analyzed to develop a taxonomy and relationships among the causal factors. Then along with other data that contributed to past incidents will be used to develop the artificial neural network (ANN) model as a predictive model. The overall objective is to develop a knowledge based predictive model to assess pipeline failure through:

a. Learning about causes behind pipeline failure: Conducting root cause analysis of past incidents to identify those factors that have to potential to contribute to failure. The findings are to be specific to the extent that they can be applied into a predictive model.
b. Implementation of learning to predict failure: Utilizing the learnings about contributing factors behind pipeline failure to develop a predictive model based on artificial neural networks that monitors current existing conditions to determine dynamic failure probability of a pipeline

2. Analysis in the Quarter

This quarter primarily focused on analysis of incident data, development of natural language processing technique, and development of an artificial neural network model. A workflow that have been used for the NLP analysis is shown in the Figure below.

Analysis of causal factors: The incident causes and sub-causes from pipeline incident data sources *e.g.,* the Pipeline and Hazardous Material Safety Administration (PHMSA), Canada National Energy Board (NEB) and European Gas Pipeline Incident Data Group (EGIG) are compared. The study considered detail analysis of causal factors of the PHMSA incident data (3,616 incident counts) on hazardous liquid (HL) pipeline for the period of 2010-2019.

Development of natural language processing (NLP) model: In the previous quarters, we preprocessed incident descriptions of HL incident data from 2010-2019 for NLP model and applied K-means and Hierarchical clustering techniques. In this quarter, we used NLP code to identify the relevant sections from a number incident investigation reports (about 50) available at PHMSA and NTSB websites and to preprocess the text from the reports. We have been trying to use topic modeling algorithm to extract useful information form the incident investigation reports.

Development of the artificial neural network (ANN) model: An ANN model has been developed for corrosion failure. The corrosion model has been targeted because there are a lot of corrosion model to compare and validate. Moreover, the number of input data are much more than other failure causes. The ANN model will be expanded include other failure causes.

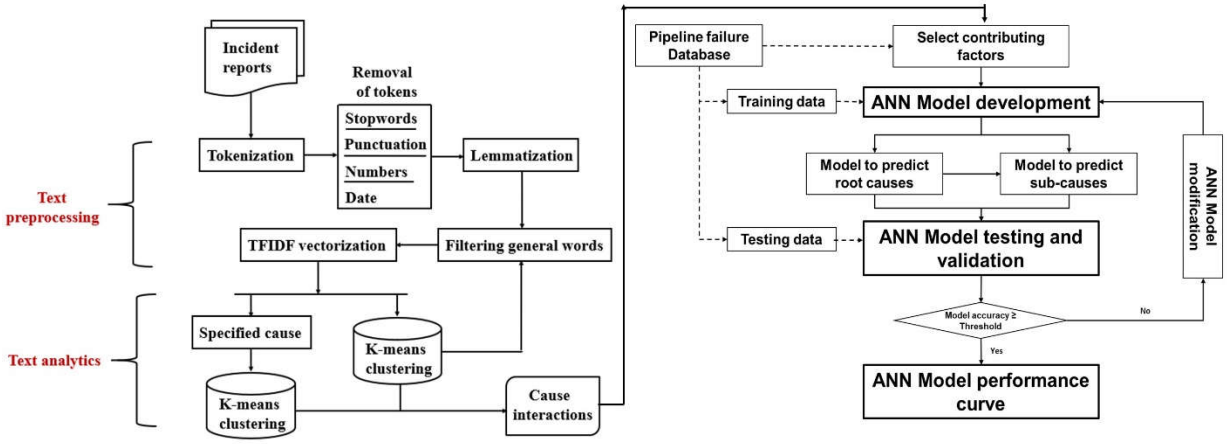

Figure Development of incident analysis tools

3. Results and Discussions

Analysis of causal factors: We have published one article on this topic in a peer-reviewed journal "Process Safety and Environmental Protection". The article has been attached. The article recommends the analysis of the causal factors contributing to the pipeline incidents should be conducted in a holistic view. The effect of one causal factor on the pipeline failure without considering the effect of other factors can be misleading. We need to be cautious not to interpret the association among the factors as causation. The article also reported how NEB is collecting underlying factors in addition to the contributing factors, which might be very useful for developing a probabilistic causal model for pipeline failure.

Development of natural language processing (NLP) model: After preprocessing K-mean clustering algorithm has been employed on incident description that involved corrosion failure (752 incident counts in HL 2010-2019). The analysis has been extended with other clustering techniques such as Hierarchical clustering technique. A manuscript has been submitted to Hazards30 conference.

Development of the artificial neural network (ANN) model: The results from the ANN model developed for corrosion failure will be presented in the MKOPSC International Symposium. An article is being drafted including the results.

Publications: Several abstracts have been submitted to conferences and symposiums and all of them have been accepted. Unfortunately, the article submitted to IPC2020 has been retracted due to current COVID-19 condition.

- Syeda Z. Halim, Mengxi Yu, Harold Escobar, and Noor Quddus, Towards a causal model from pipeline incident data analysis, Journal of Process Safety and Environmental Protection. (Published)
- Guanyang Liu, Mason Boyd, and Noor Quddus, Extracting Causal Relations from Incident Reports: A Natural Language Processing and Topic Modeling, 2020 Spring Meeting & 16th Global Congress on Process Safety (accepted for both presentation and poster)
- Pallavi Kumari and Noor Quddus, Causation Analysis of Pipeline Incidents Using Artificial Neural Network, 2020 Spring Meeting & 16th Global Congress on Process Safety (accepted for

presentation)

- Guanyang Liu, Mason Boyd, and Noor Quddus, Analysis of Pipeline Incident Data and Investigation Reports Using Natural Language Processing (NLP), paper submitted to Hazards30, 18-20 May, Manchester, UK (accepted).
- Guanyang Liu, Mason Boyd, Pallavi Kumari, Syeda Zohra Halim, and Noor Quddus, An Intelligent Learning Framework for Analysis of Pipeline Incident Investigation Reports, Submitted to Process Safety International Symposium 2020, College Station, October 2020. (accepted for presentation)

4. Future work

Future work will be continued on the following

- The NLP model development will be continued and applied to incident investigation reports available at PHMSA and NTSB and to identify the causal factors, background factors and underlying factors. NLP techniques such topic modeling and others will be explored for best extraction of information using the dataset.
- ANN model development will be extended to include other failure causes and predict the future incident.