**CP2-3: CSCI 8715**

**Tentative Title:** Predicting Pipeline Failure Risk using Pipeline Incident Reports and GIS Imagery data for Pipeline, Terrain and Population Characteristics.

**Submitted by:** Bhupinder Juneja (junej003), Sagar Tamang (taman011) and Prof. Sehwon Kang (external collaborator).

**2. Research Abstract**

Pipeline safety is vital because pipeline incidents threaten human health, biological life, and the environment. One single pipeline incident can result in a significant consequence. For instance, the cleanup expense of the Kalamazoo River incident was estimated at 1.21 billion dollars. While the pipeline industry faces considerable pressure of monitoring, scrutiny, and sanction and the stakeholders (e.g., pipeline operators, Department of Transportation) have increased investment in complying standards, preventing pipeline incidents is still very challenging because it is difficult to predict the risk of the pipeline system. First, the total pipeline mileage in the United States is more than 2 million miles and pipeline incidents are low-probability high-consequence events. Second, traditional statistical approaches cannot deal with high dimensional data from geographical characteristics (e.g. hydrography, waterway, land cover, pipeline network shape). In this respect, our study is focused on developing and validating the performance of predictive models of future pipeline risk severity and likelihood by analyzing multiple raster images created from geographical information (e.g., hydrography, population density, protected areas, and, commercially navigable waterway).

**Research Questions**

1) Descriptive analysis

- What is the impact of geographical characteristics (e.g. hydrography, population) on pipeline risk: probability and consequence?

2) Predictive analysis

- Given data, predict pipeline incident frequency. Which data/factors will exhibit predictive power?

- Given data and incidents, predict the pipeline incident consequence. Which data/factors will exhibit predictive power?

3) Prescriptive analysis

- Is it possible to suggest an alternative pipeline route to reduce short term and long term risks and consequences?

**Data Sources Identified:**

1) Pipeline Incident Data (2010 – 2017): CSV

- Time: incident time, incident identified time, reported time

- Consequence:

Volume: released commodity type, intended release, unintended release, recovered,

Cost and other: fatality, injury, property damage, env. cost, total cost, commodity lost cost

Ecological: wildlife impact (Y/N), soil contamination, water contamination

- Location: incident Longitude/Latitude, onshore/offshore indicator,

Onshore: area type (above ground, underground, transition, tank), sub area type (e.g. under soil), depth of cover

Offshore: water depth, area type (e.g. below water, shoreline)

- Pipeline Operator: id, name, address, (incident – pipe facility name, segment name)

- Pipeline characteristics: intra/inter state, system part involved incidents, item involved incidents, pipeline thickness, pipeline diameter, seam type, material type, installed year, pipeline pressure

- incident cause: external corrosion, internal corrosion, natural force (earth movement, floods, lightning, temperature, high winds, other), excavation damage, other (car, boat, fishing, …), material failure, equipment failure, incorrect operation

2) Pipeline system: image/GIS

- Petroleum product terminal: All operable bulk petroleum product terminals located in the 50 States

company name, city, state, longitude & latitude

- border crossing liquids: Border crossings of liquids pipelines. A crossing point represents one or more pipelines

company name, city, state, number of pipeline, diameter, longitude & latitude

- Petroleum product pipeline: operator name, pipeline name

- HGL pipeline: operator name, pipeline name

- Crude oil pipeline: operator name, pipeline name

3) Population: image/GIS

- high populated area: a population density of at least 1,000 people per square mile

- other populated area: area contains a concentrated population, such as an incorporated or unincorporated city, town, village, or other designated residential or commercial area

4) Waterway: image/GIS

- Commercially Navigable Waterway: name, depth, length and shape

5) Protected Area: Shape - nation's official inventory of protected areas in the standard PAD-US framework (biodiversity protection): image/GIS

manager agency, unit type, bio diversity gap code, shape, year and others

6) Hydrography: image/GIS

- water-related entities, such as industrial discharges, drinking water supplies, fish habitat areas, wild and scenic rivers. upstream/downstream relationships of these water-related entities--and any associated information about them--can be analyzed using software tools ranging from spreadsheets to geographic information systems

**References:**

1. Okabe, Sugihara, Sugihara, Kōkichi, and Ebrary, Inc. Spatial Analysis along Networks: Statistical and Computational Methods. Statistics in Practice. Hoboken, N.J.: Wiley, 2012.
2. Barthélemy, Marc. "Spatial Networks." Physics Reports 499, no. 1-3 (2011): 1-101.
3. A. Ouadah, "Pipeline Defects Risk Assessment Using Machine Learning and Analytical Hierarchy Process," 2018 International Conference on Applied Smart Systems (ICASS), Medea, Algeria, 2018, pp. 1-6.
4. [Pipeline condition prediction using Neural Network Models](https://ascelibrary.org/doi/abs/10.1061/40800(180)61)
5. [Safety Assessment of oil and gas pipelines using satellite information](https://www.sciencedirect.com/science/article/pii/S2452321619306304)
6. Baghdadi, A.H.A, and H.A Mansy. "A Mathematical Model for Leak Location in Pipelines." Applied Mathematical Modelling 12, no. 1 (1988): 25-30.