CAAP Quarterly Report

Date of Report: *July 7th, 2020*

Contract Number: *693JK318500010CAAP*

Prepared for: *U.S. DOT Pipeline and Hazardous Materials Safety Administration*

Project Title: Brain-Inspired Learning Framework to Bridging Information, Uncertainty and Human-Machine Decision-Making for Decoding Variance in Pipeline Computational Models

Prepared by: *North Dakota State University*

Contact Information: *Ms. Zi Zhang, PhD student, Email: zi.zhang@ndsu.edu@ndsu.edu, Phone: 701-231-7204; Dr. Hong Pan, Postdoc, Email: hong.pan@ndsu.edu; Mr. Matthew Pearson, M.S. student, Email: matthew.pearson@ndsu.edu, Phone: 701-231-7204; Dr. Zhibin Lin, Email:zhibin.lin@ndsu.edu, Phone: 717-231-7204*

For quarterly period ending: *July 7th, 2020*

Business and Activity Section

(a) Generated Commitments

Top journal paper published: a journal paper, entitled "*Machine learning-enriched Lamb wave approaches for automated damage detection*" was published in a top Journal - Sensors (Impact factor=3.031). PhD student Zi Zhang who mainly takes charge of this research was the first author.

Conference paper accepted under virtual presentation duo to Covid-19 situation: two conference papers, entitled "*Corrosion-induced damage detection and conditional assessment for metallic civil structures using machine learning approaches*" and "*Conditional assessment of large-scale infrastructure systems using deep learning approaches*", were accepted as conference papers and presentation under virtual presentation duo to Covid-19 situation, *2020 SPIE Smart Structures and Nondestructive Evaluation, April 26-30,* Anaheim, California, USA.

(b) Status Update of Past Quarter Activities

(c) Cost share activity

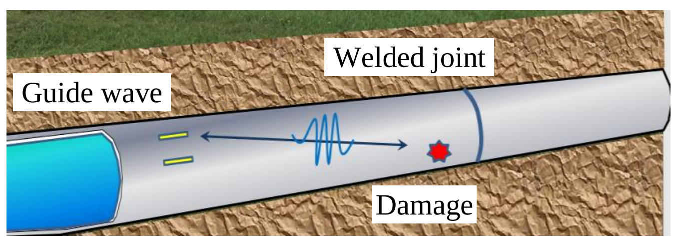
Cost share was from the graduate students' tuition waiver.

(d) Summary of detailed work for Tasks 4-6

This task aimed to address the challenge associated to measurement noise, operational, and nonlinearity due to material discontinuity.

As for large volumes of ultrasonic data, numerous uncertainties and complex guided wave propagation in oil/gas pipeline, as schematically in **Fig. 1**, feature extraction may cause the limitation. However, deep learning methods, such as convolutional neural networks (CNN), was used herein to enhance information extraction and better classify structural uncertainty of from data in pipeline associated with a high levels of variances, including measurement noise, nonlinearity and other uncertainty. Clearly, oil/gas pipeline structures are often exposed to complex environment with high levels of uncertainty, as schematically shown in **Fig. 1b**. As a result, Lamb wave signals collected from complex structural systems in fields could be highly affected by structural uncertainty, which in turn affects the effectiveness of the methods for engineering applications. We discussed the simulation model of the pipeline and several states were designed to detect the damage using CNN algorithm. As stated in last report, the framework of this study was shown in **Fig. 2**. The study was presented herein to address the effectiveness of the proposed deep learning methods when handling structural uncertainty due to noise level and material discontinuity from weldment that pipe engineers often face with in field.

(a) Pipes used for oil/gas transmission line



(b) Schematics of guided wave along a pipeline

Fig. 1 Data collected from guided wave along a pipeline

The merits of using the proposed learning framework over conventional physics-based signal process were mainly on:

(a) Handling nonlinear and high-dimensional features; physics-based features such as amplitude, phase change, and correlation coefficient, which are often used explicitly for determining damage level and size, could be insensitive to defects in some cases when facing with complexity of guided wave multimodal interaction, noise or other interference. Differently, deep learning could automatically extract sensitive features related to structural and material discontinuity, with less physical representation.

(b) Tackling more structural complexity with less physical restraints; Guided wave exhibits non-stationary and nonlinear behavior, experiencing complex dispersion and coherent multi-mode interaction. Different to physics-based methods that attempt decomposition of mixed modes for signal process, the machine learning could extract sensitive damage features, with less or without such physical restraints. As a result, with representative data, the machine learning could provide better damage detection with minimized explicit formation that physics-based methods highly rely on.

(c) Uncovering structural uncertainty; Consider that oil/gas pipelines are often exposed to high levels of uncertainty, structural uncertainty is one of challenges for physics-based methods. The cases were designed in this study to address this challenge and demonstrate the effectiveness of the proposed learning framework under structural uncertainty due to noise level and material discontinuity from weldment. The findings were expected to provide new vision using machine learning methods for pipeline engineering applications.

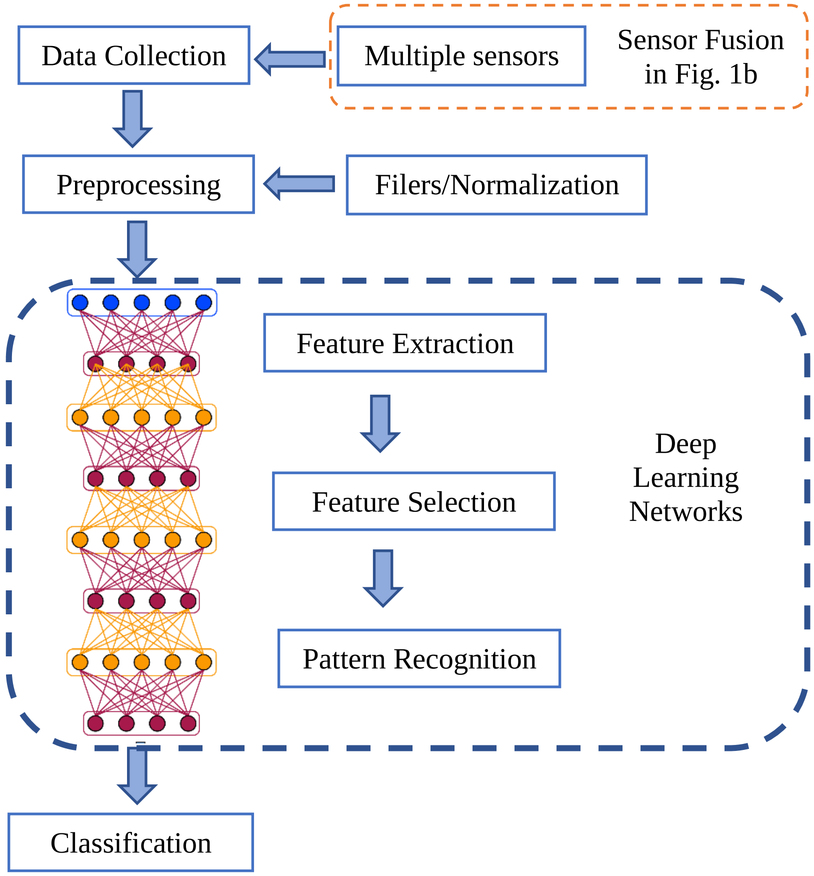


Fig. 2 Framework in this study

Oil/gas pipeline was simulated using 3D FE modeling through COMSOL. The prototype of a steel pipeline was selected from the literature **Error! Reference source not found.**, where its dimension is 76-mm in outside diameter and 4 mm in wall thickness, and with a length of 2000 mm.

Fig. 3 indicated the details of guided waves propagated and diffracted through the whole span of pipeline during different periods. The excitation wave was input from the left side and traveled to the right. When the signal interacted with the weldment at 1.8E-4 s, part of the wave was return, and the rest was continually moving forward. At 3.2E-4 s, the initial wave was interfaced by the notch-shaped damage, and on the other side, the echoed wave was arrived at the left boundary.

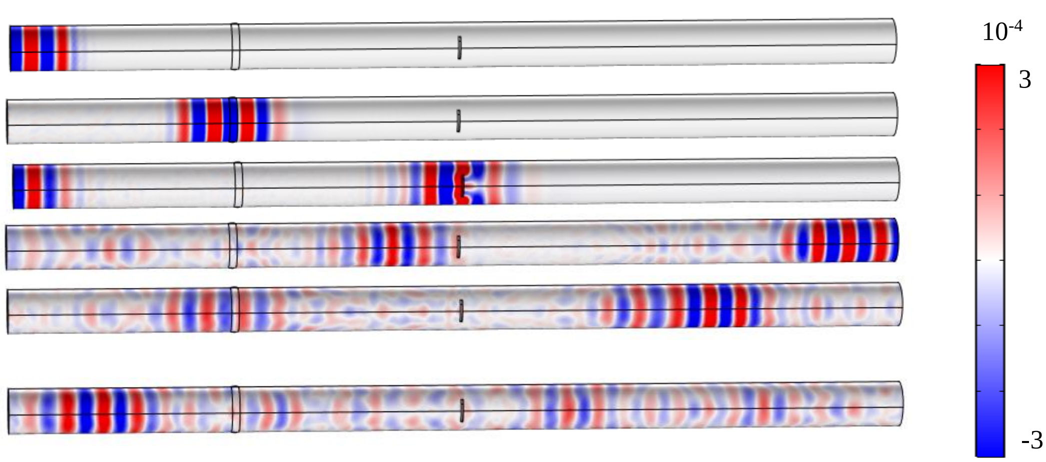


Fig. 3 Wave propagation through the whole span

Convolutional neural network was involved to identify the damage in the pipe by guided wave. As illustrated in **Fig. 4**, the learning framework was consisted by three main parts, including dataset collected by the simulation, Training CNN models by training set and modifying by validation set, and predicting the testing data formed by CNN models.

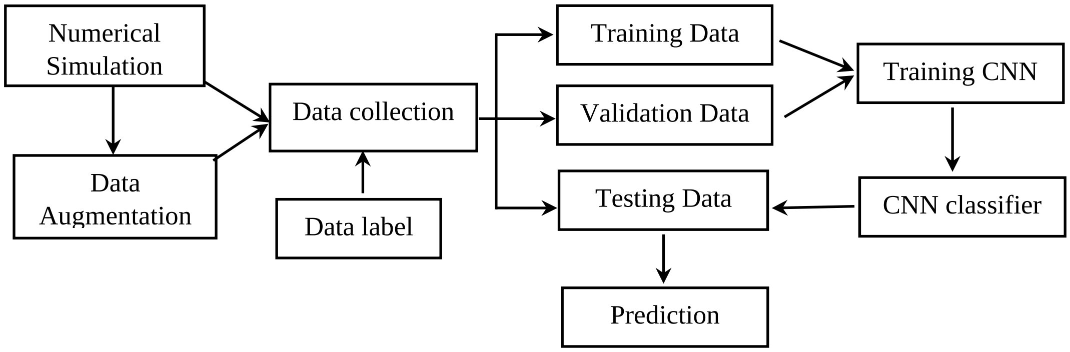
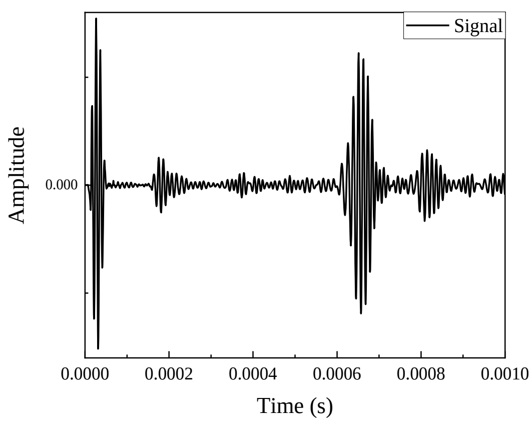
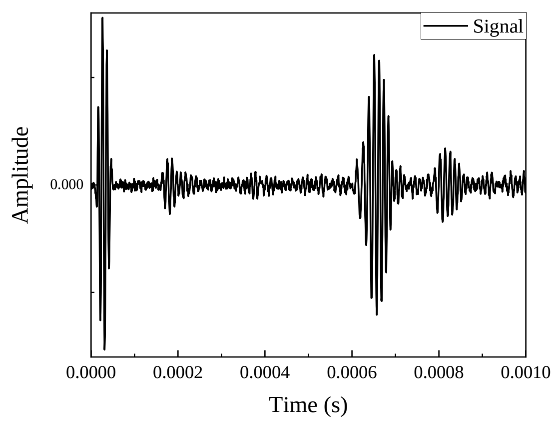


Fig. 4 Flowchart for classification utilizing CNN

The input data is significant for training the network. Adding noise into the clear data can increase the uncertainty and complexity. However, it is hard to classify the data when they covered by noise in high level. The signals under different noise levels were shown in **Fig. 5**. The noise with 120 dB and 100 dB were not affect the main information of the signal. When it increased to 80 dB, some wave packages with low amplitude were devoured by the noise, but the main packages can still be identified.



(a) Original signal

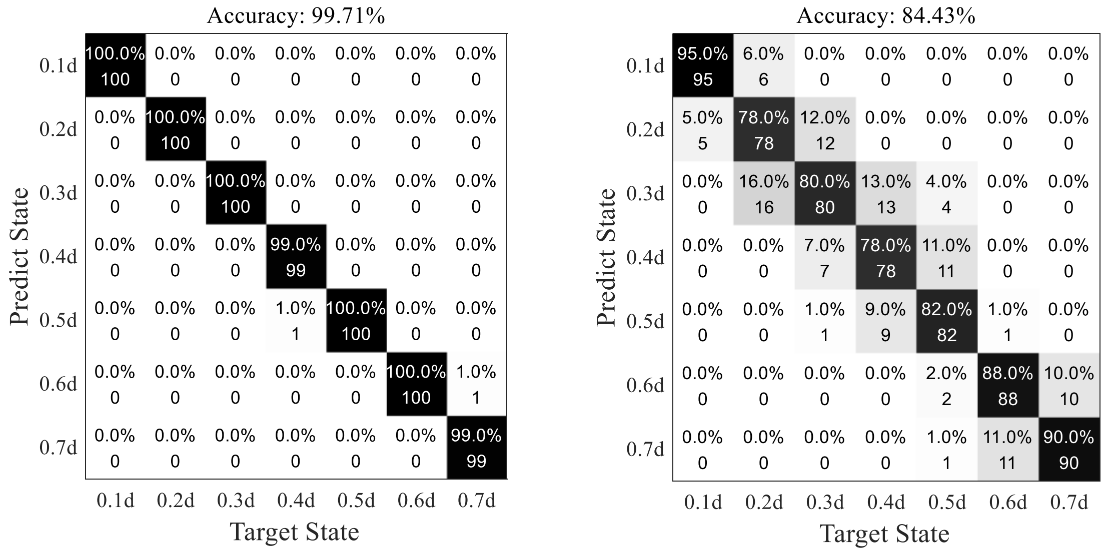


(b) SNR = 100 dB

Fig. 5 Impacts of noise levels on signals

4.1.1 Testing results of damage size and depth affected by noise level

Fig. 6 illustrated the classification results of the test set, distinguishing these signals into 7 different damage size. The data set could be completely classified into the correct categories when SNR was larger than 70dB. Although the training set was 100% classified by CNN at 70 dB, some errors still occurred when classifying the testing set. Specifically, one signal from the damage sized 0.4*D_{out} in State #10 was misled into State #5 (0.5*D_{out}), and one belonging to State #12 was misclassified into State #11. When SNR decreased to 70 dB, the accuracy of the test was reduced to 84.43%. **Fig. 6(b)** showed the confusion matrix of the prediction. Briefly, misleading judgments occurred in adjacent groups. For instance, 6% of the signals representing the 0.2*D_{out}-long damage was classify into 0.1* D_{out} class, and 16% was misjudged into 0.3* D_{out} class.



(a) SNR = 70 dB (b) SNR = 60 dB

Fig. 6 Testing results of damage size

Task 5: Heterogeneous Data Sets

This task aimed to address the challenge due to various heterogeneous data sets and unveil the fundamental nature of various data types.

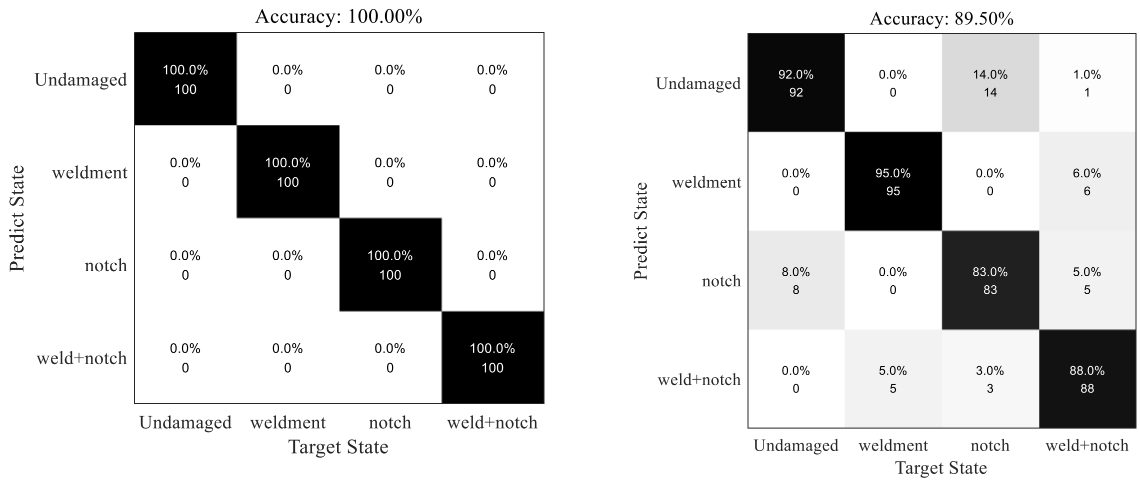
Sub-Task 5.1. Mixed fusion method for heterogeneous data sets

First attempt was to category the multiple type and understand how to mitigate the influence of different contamination. To accomplish this task, we used a mixed data fusion strategy to clear and assemble data. This mixed data fusion process including sensor level fusion and central-level fusion. The final result of this data fusion process is united to three different formats based on the original signal's format and the application purpose.

Sub-Task 5.2. Data classification between image and sensory data

The second attempt was made using common images and sensory data. The results of the damage detection were 4 different signals from receivers located circumferentially around the pipe. Considering the interrelationship between signals from different locations, four signals were combined as a matrix input into the deep learning.

Fig. 7 represented the confusion matrix of the testing data in SNR equal to 70 dB and 60 dB respectively. When the noise level is lower than 70 dB, the accuracy of the prediction for the testing data was 100%. However, in 60 dB condition, the accuracy of the identification was 89.5%.



(a) SNR = 70 dB (b) SNR = 60 dB

Fig. 7 Confusion matrix for time-series data classifcation

6.1.1 Deep Bayesian Belief Network (DBBN)

Learning architecture of stochastic binary variables between layers ***v*** and ***h*** through the energy-based method using Boltzmann machine (Hinton et al. 2006; Hinton et al. 2012):

$$E(\boldsymbol{v},\boldsymbol{h}) = -\sum a_i v_i - \sum b_j h_j - \sum h'_j W v_i - \sum v'_i U v_i - \sum h'_j V h_j \quad (1)$$

where, a and b is the biases of the stochastic variables ***v*** and ***h***, respectively; W, U, V are the weights of each connection; and the joint states of v' and v *(or* h' and h) denote the adjacent connection of the variables within a layer. As illustrated in **Fig. 50**, the deep Bayesian belief network (Zhao et al., 2015; Chaturvedi et al., 2016) is the multiple layers neural networks. The model is effective to perform top-down and also the opposite bottom-up generative weights, which allows using back-propagation for fine-tuning for optimized discrimination/regression.

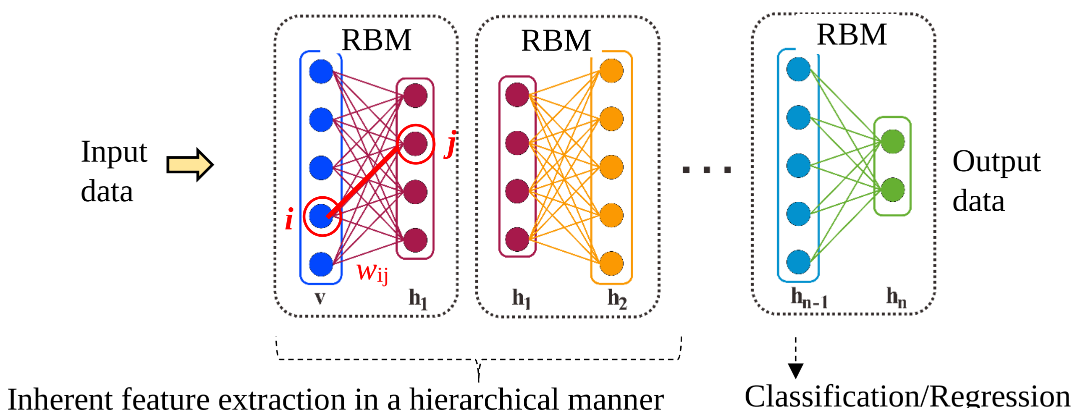


Fig. 8 Architecture of the DBBN

The DBBN are constructed using multiple restricted Boltzmann machines (RBMs) (Hinton et al. 2006; Hinton et al. 2012; Zhao et al., 2015). As schematically illustrated in **Fig. 8**, the architecture of the DBBN consists of undirected multiple levels of the RBMs, where the hierarchical architecture could allow automation of feature extraction from lower to upper layers, and the final layer could be constructed using different activation function for either classification or logistic regression of interest.

6.1.2 Concept of the RBM and its architectures

To avoid the complexity and difficulty in determining parameters in Boltzmann machine, as shown in Eqn. (5b), the RBM is developed by an undirected graphical machine without visible-visible or hidden-hidden connections (Mohamed et al. 2012). Take one unit of the RBM (e.g., RBM_1 in **Fig. 8**) as an example, the visible variable v_i, and the hidden variable, h_j, are connected and assigned by a weight, w_{ij}. The probability to the joint states of the visible and hidden vector is defined by the energy-based function (Hinton et al. 2006; Hinton et al. 2012)

$$p(\boldsymbol{v},\boldsymbol{h}) = \frac{e^{-E(\boldsymbol{v},\boldsymbol{h})}}{\sum_{i=1}^{V}\sum_{j=1}^{H} e^{-E(\boldsymbol{v},\boldsymbol{h})}} \tag{2}$$

where, $\sum$ is the summation over all visible and hidden variables and the $E()$ is the energy-based function. Consider structural data of interest could be binary data, such as black or white color in image recognition, or be more complex sensory information, three different data types (either in visible or hidden layer) are defined herein by (Hinton et al. 2010):

(e) Description of any Problems/Challenges

No problems are experienced during this report period

(f) Planned Activities for the Next Quarter

The planned activities for the next quarter are listed below:

- Modeling variance experienced from structural uncertainties, including experimental tests.
- Modeling variance from heterogeneous data sets.