**Training Machine Learning Models to Assist Guided Wave NDE Inspectors with Flaw Detection and Characterization**

**Kayte M Denslow1, William S. Rosenthal1, William R. Johns1, and M. Subramanian1**

1Pacific Northwest National Laboratory

902 Battelle Blvd.

Richland, Washington 99354

(509) 375-2232; Kayte.Denslow@pnnl.gov

ABSTRACT

Two robotic ultrasonic guided wave inspection technologies, RAVIS and RREVIS, have been developed to overcome a long-standing inspection challenge at the Hanford site – the volumetric inspection of physically inaccessible 3/8- to 7/8-in. thick carbon steel floor plates of the primary (inner) tanks in buried double-shell tank systems. Despite the use of the non-dispersive fundamental shear-horizontal guided wave mode by RAVIS and RREVIS to simplify signal and image data, other noise, artifacts, or reflections from other wave modes created during mode conversion can still complicate the signal/image data. As a result, NDE inspectors/analysts are challenged with distinguishing between signal or image features associated with flaws and features associated with artifacts or other wave modes reflections. The inability to accurately deduce flaw type and size by applying standard amplitude and time thresholding to a flaw indication is another challenge that impedes an analyst’s ability to accurately classify and estimate these flaw characteristics.

The Hanford tank operations contractor Washington River Protection Solutions has teamed with Pacific Northwest National Laboratory, Guidedwave, and the Electric Power Research Institute to train a machine learning model to overcome the data interpretation and analysis challenges for RAVIS. The model selected is a Siamese Neural Network, which is a classification model designed to perform well given relatively small amounts of training data. The model is being trained to learn how to correlate ultrasonic guided wave transducer signal features with physical flaw and plate features so the trained model can eventually be presented with new tank inspection signal data and analyze its signal features to (1) distinguish between noise and reflections associated with welds and flaws and (2) classify the flaw type and size with which the flaw-related signals correspond. The trained model is intended to be a companion to RAVIS inspectors/analysts to help identify which indications in future images of tank floor plates should be considered flaws and decide whether a flaw meets U.S. Department of Energy allowable flaw size criteria. The same model framework could be used to develop a companion analysis tool for RREVIS inspectors too. The approach to training the Siamese Neural Network and the progress made toward training it are presented here.

**Keywords:** Hanford, double-shell tanks, DST, nuclear, radioactive, ultrasound, ultrasonic, guided wave, shear horizontal wave mode, SH0, volumetric, non-destructive evaluation, NDE, machine learning, neural network, Siamese neural network, RAVIS, RREVIS, EMAT, GWPA

**PRIMARY TANK FLOOR INSPECTION IN HANFORD DOUBLE-SHELL TANKS**

Two robotic inspection technologies – the Robotic Air-slot Volumetric Inspection System (RAVIS) developed by Guidedwave and Eddyfi Technologies and the Robotic Remote Electromagnetic-acoustic-transducer Volumetric Inspection System (RREVIS) developed by Southwest Research Institute (SwRI) – have been engineered for the tank operations contractor, Washington River Protection Solutions (WRPS) [1]. Their purpose is to overcome a long-standing inspection challenge at the Hanford site – the volumetric inspection of physically inaccessible 3/8- to 7/8-in. thick carbon steel floor plates of the primary (inner) tanks in buried double-shell tank (DST) systems. The 1-1.5-million-capacity, 75-ft. diameter primary tanks must be remotely inspected from outside the tanks while they are in service (hold radioactive waste) to detect corrosion pits, general corrosion (wall thinning), and intergranular stress-corrosion cracks that may originate from the internal and/or external plate surfaces.

RAVIS and RREVIS will inspect the welded carbon steel primary tank floor plates using low-frequency ultrasonic guided waves emitted from transducers placed in remote standoff positions on physically accessible parts of the tank exterior – either the narrow strips of tank floor plate directly over refractory pad air vents (RAVIS) or the lower tank sidewall just above the lower knuckle plates (RREVIS). The long-range inspections from standoff transducer positions enable large swaths of inaccessible (and accessible) primary tank floor plates to be inspected in minutes. Table 1 provides the flaws they must be able to detect and locate (**NDE Objectives 1-2**) and characterize (**NDE Objective 3**).

Table 1: Flaw acceptance criteria and actions established for Hanford waste tanks [2-4].

|  |  |  |  |
| --- | --- | --- | --- |
| **Degradation/Flaw Type** | **Reportable Level Valuesa**  **(for Documentation)** | **DOE Acceptance Criteria**  **Action Level Valuesa** | **Corrective Action Level** |
| Pit (pitting corrosion) | 25% t | 50% t | >75% t |
| Crack/weld seam opening | Any linear indication greater than 6 in. long and 0.1 in. deep | If >12 in. long, 20% t  If <12 in. long, 50% t |
| Wall thinning (general corrosion) | 10% t | 20% t |
| t = thickness of plate | | | |

RAVIS and RREVIS use different transducer designs – a 150-165 kHz piezoelectric guided wave phased array transducer (GWPA) and a 43-57 kHz electromagnetic acoustic transducer (EMAT), respectively – but both rely on generating, propagating, and receiving the fundamental shear-horizontal guided wave mode (SH0) to inspect the plates while simplifying signal and image data. However, undesirable high-amplitude ultrasonic reflections associated with noise/artifacts or other wave modes created during mode conversion can be comingled with desired SH0 reflections from flaws that complicate the signal/image data. Consequently, NDE inspectors/analysts are challenged with distinguishing between signal/image features associated with flaws and features associated with artifacts or other wave modes reflections (**NDE Objectives 1-2**). The inability to accurately deduce flaw type and size by applying standard amplitude and time thresholding to a flaw indication is a ubiquitous challenge for guided wave NDE, which impedes an analyst’s ability to accurately classify and estimate flaw dimensions (**NDE Objective 3**).

**Machine Learning Model to Support GWPA Transducer Data Analysts with Flaw Detection and Characterization**

The RAVIS’ GWPA transducer was the first to undergo qualification testing on a 24 ft. wide by 48 ft. long full-scale mock-up of 1/8th of a primary tank floor that was designed for blind NDE performance demonstrations to assess NDE reliability and qualify the transducer for inspections. The performance demonstrations revealed the GWPA images of coherent SH0 energy reflections from welds and flaws are contaminated with (1) coherent reflections of the dispersive first higher-order shear-horizontal wave mode (SH1) and (2) noise/artifacts caused by transducer ringing due to inadequate transducer coupling conditions and coincident reflections from multiple reflectors that are equidistant from the transducer. GWPA analysts used conservative signal metrics to discriminate between flaws and noise/artifacts, but the conservatism negatively affected the probability of detection (POD) by 15-20%. Reducing the SH1 signals and noise/artifacts through GWPA transducer design has been attempted, but the size constraints placed on the GWPA transducer have made it challenging to get the 26 piezoelectric elements to a size and resonance frequency that would predominantly excite the SH0 wave mode. The nominal 165 kHz frequency of the transducer is above the cutoff frequency for SH1 wave mode and consequently the SH0 signals are accompanied by SH1 signals. Advanced analysis must be used to help discriminate between flaws and noise/artifacts.

The data collected on the primary tank floor qualification mock-up to qualify the GWPA transducer are being leveraged to begin developing an advanced analysis tool for GWPA transducer data analysts to help them discriminate between flaws and noise/artifacts, and subsequently classify a flaw by its type and size. The analysis tool is based on supervised machine learning since it offers a way to (1) uncover the measurable waveform/signal features that uniquely correlate with flaws and flaw type to support accurate flaw/noise/artifact sorting, and (2) build complex multi-variable correlations between waveform/signal features and physical flaw features that can be relied upon to identify the type and size of plate flaws based on waveform/signal features in the plate inspection data. Specifically, a Siamese neural network (SNN) has been selected for the machine learning task. It is a classification model that learns how to tell the difference between different classes of inputs and is designed to perform well given relatively small amounts of training data, provided the number of flaw type classes is not infinite [5]. When the input classes for the SNN are transducer signal features that correspond to a few canonical flaw types, then it becomes capable of sorting any new transducer signal data according to the closest available flaw type. The output of the SNN also includes the relative probability that each flaw type is a match, which can be used to quantify the uncertainty in the predictions.

***Training Progress***

Each NDE performance demonstration performed on the primary tank floor plate mock-up with the GWPA transducer provided a new batch of data for the SNN model. A total of four NDE performance demonstrations were required to test the transducer on a statistically significant set of flaw scenarios and flaw-free (null) scenarios, where the condition of the primary tank floor plate mock-up changed between each NDE performance demonstration. Thousands of GWPA scans have been performed that collectively contain tens of thousands of raw ultrasonic time-series signals, which are the data that are useful for training a machine learning model how to (1) distinguish between flaws and noise/artifacts to detect and locate flaws (**NDE Objectives 1-2**) and (2) classify flaws by their type and size (**NDE Objective 3**). Only one NDE performance demonstration was performed per year and SNN training occurs the year after the data are collected. The results from machine learning model training shared here are those which were generated with an SNN model trained on GWPA transducer data collected through 2022 only. The SNN model’s training data set is being updated with GWPA transducer collected in 2023, whereupon the SNN model will be retrained and its flaw vs. null discrimination performance and flaw class prediction performance will be re-evaluated.

***Training Process***

The GWPA transducer performed a 360-degree scan every 12 inches along mock air vent paths of the 24 ft. wide, 48 ft. long primary tank floor plate mock-up (see Figure 1). A multitude of scans from different offset distances and angles were collected relative an “area of interest” in the mock-up. The “proximal sensor locations” used in the training data set for the SNN model were only those within 44 inches of a flawed or flaw-free (null) area of interest, as this is the scan radius over which the transducer data are of highest quality. A given area of interest on the plate may have more signal data associated with it than other areas of interest depending on the quantity of air vent paths around each area of interest in the mock-up.

A picture containing outdoor, blue, dock, line

Description automatically generated 

**Figure 1: Photos from the 2023 NDE performance demonstration. (a) The GWPA transducer being coupled to the topside of the primary tank floor test mock-up in an air vent path that exposes strips of the plate for transducer coupling. (b) A 360-degree scan displayed in real time by the Guidedwave data acquisition system.**

To curate the GWPA transducer signal data for machine learning model training, it is first organized relative to each null (flaw-free) or flawed area of interest selected for SNN model training [6-7]. The waveform(s) of interest in the signal data that correspond with the null or flawed area of interest are subsequently analyzed to reduce the digitized waveforms to a set of signal metric values (peak amplitude, peak width, energy, peak frequency, frequency width, phase, and noise amplitude). The physical areas of interest in the mock-up are also analyzed to reduce them to a set of physical metric values (max flaw depth, flaw area, length of major axis of flaw, roundness, compactness, skewness, and eccentricity). Only eight proximal sensor locations within 44 inches of an area of interest were used at a time during model training, though a given area of interest may have multiple sets of eight different proximal sensor locations if GWPA transducer scans were collected at more than eight standoff distances and/or approach angles around the area of interest. The number of potential flaw classes or outputs in the SNN model was limited to eight to limit the algorithmic complexity of the SNN classification model to be on par with the amount of data in the dataset. One flaw class=null (no flaw) and the other seven flaw classes represent different combinations of values for a given set of physical characteristics (e.g., maximum flaw depth, flaw area, flaw eccentricity).

***SNN Flaw Classification Performance for GWPA Transducer Data***

An array of GWPA transducer signal data for flawed plate areas of interest only (no null locations yet) was fed to the model first, and the model’s prediction of flaw class was recorded during model training. Each one of these cycles is termed an epoch. The discrepancy between the model’s prediction and the correct answer is “back-propagated,” i.e., used to adjust the parameters of the model so it would be more accurate for the next epoch. This process was repeated for approximately 100K epochs, although often the training converged on a “best” answer for the model parameters in about 30K epochs. The model’s prediction accuracy could fluctuate near the end of the training period, so the accuracy statistics were monitored to assess convergence. This training process was repeated with a fresh initialization of the model (starting with randomized parameters) a total of five times, over which the average prediction accuracy was 97.3%.

***SNN Flaw vs. Null Discrimination Performance for GWPA Transducer Data***

The full array of GWPA transducer signal data for flawed plate areas of interests and flaw-free/null areas of interest were fed to the model next to assess the model’s ability to discriminate between flawed and unflawed scenarios, i.e., to evaluate how well it could classify a flaw scenario as any one of any of the seven possible flaw classes (even if it was the wrong flaw class) and how well it could classify a flaw-free scenario as the null class. False positives, or predictions of significant flaws where there weren’t any, are important errors to minimize through the design of the SNN model framework. A set of 400 areas of interest (combination of flaw and null locations) or “trials” were used to determine how accurately the SNN model could classify the trials as flaws or nulls. Each trial in the experiment included a set of GPWA transducer signal data associated with eight different offset distances and/or approach angles within a 44-in. radius of an area of interest (each set of eight was randomly chosen per epoch). The correct call rate (true positives and true negatives combined) was 87.5%. The incorrect call rate (false positives and false negatives combined) was 12.5%, where false positives (incorrectly predicting a flaw was present, i.e., a false alarm) were 7.5% of the calls and false negatives (incorrectly predicting a null, i.e., missing a flaw) were the other 5.0% of the calls.

**Ongoing and Future Work**

The GWPA transducer signal data from the 2023 NDE performance demonstration is being analyzed and curated with physical feature data on corresponding areas of interest in the mock-up. The data will be added to the SNN model’s training data set and the final round of SNN model training on empirical data will be performed. The updated SNN model will be re-tested on flaw and null trials to re-evaluate its flaw vs. null discrimination performance and flaw classification performance. The results will determine if the empirical data from the NDE performance demonstrations are sufficient to support robust flaw detection and location (**NDE Objectives 1-2**), and whether the estimated 15-20% increase in POD over that obtained by human analysis is possible. The results will also determine if the empirical data are sufficient to support robust classification of flaws by flaw class (part of **NDE Objective 3**) and the importance of generating and validating simulated and/or synthetic GWPA transducer signal data to the training data set to achieve flaw characterization (**NDE Objective 3**). The goal for the model is to accurately classify a flaw by type (corrosion pit, general corrosion, or crack) and estimate flaw depth with respectable accuracy (e.g., ±10-15%) to provide the information needed to support leak risk assessments and tank management decisions.

The RREVIS will begin qualification testing on the primary tank floor test mock-up in the near future. If the NDE performance demonstrations reveal the EMAT signal and image data are complicated by SH1 and/or noise and other coherent reflections that negatively affect the POD, then the EMAT signal data can be leveraged to create another SNN model to create an advanced analysis tool for EMAT transducer data analysts to help them discriminate between flaws and noise/artifacts, and subsequently classify a flaw by its type and size.

**REFERENCES**

1. Hede A.L., A.R. Rodriguez, K.D. Boomer, G.E. Soon, J.R. Gunter, and K.M. Denslow, 2023, *Hanford Double-Shell Primary Tank Bottom Inspection Technology Development,* Waste Management Symposia, March 10-14, 2024, Phoenix, AZ.
2. J. R. Follett, 2017, “Leak Inspection Report for Tank 241-AY-102,” RPP-RPT-60320, Rev. 0, Washington River Protection Solutions, Richland, WA.
3. C. L. Girardot, J.R. Gunter, N.M. Young, and J.S. Garfield, 2019, “Double-shell Tank Integrity Program Plan,” RPP-7574, Rev. 6, Washington River Protection Solutions and AEM Consulting, Richland, WA.
4. K. Bandyopadhyay, et al., 1997, “Guidelines for Development of Structural Integrity Programs for DOE High-Level Waste Storage Tanks,” BNL-52527, Brookhaven National Laboratory, Upton, NY.
5. G. Koch, et al., 2015, *Siamese Neural Networks for One-shot Image Reconstruction*, 32nd International Conference on Machine Learning, Lille France, Journal of Machine Learning Research Workshop and Conference Proceedings, Vol. 37.
6. W. S. Rosenthal, K.M. Denslow, J.M Brandi-Lozano, M. Subramanian, W.R. Johns, F.C. Grogan, S.K. Cooley, K.D. Boomer, A.L. Hede, T.A. Wooley, J.R. Gunter, G.E. Soon, E.N. Nelson, C.J. Borigo, R.G. Love, M. Quarry, and L. Breon, 2023, *Next-generation Flaw Characterization in Hanford Under-tank Inspection by the Robotic Air-slot Volumetric Inspection System (RAVIS)*, Waste Management Symposia, February 26-March 2, 2023, Phoenix, AZ.
7. K.M. Denslow, W.S. Rosenthal, N. Karri, K. Kadooka, K.D. Boomer, A.L. Hede, J.R. Gunter, G.E. Soon, C.J. Borigo, R.G. Love, A.D. Reese, and M.J. Quarry, 2024, *Next-Generation Machine Learning-Assisted Robotic Ultrasonic Inspection Technology for Hanford Primary Tank Floor Plates*, Waste Management Symposia, March 10-14, 2024, Phoenix, AZ.