U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Artificial Intelligence and Machine Learning in Nondestructive Examination and In-Service Inspection Activities

Carol A. Nove: RES/DE
Stephen Cumblidge: NRR/DNRL

ACRS Full Committee Briefing
March 6, 2024

Outline

- Introduction and Background
- Research Program
  - Evaluation of commercially available automated data analysis
  - Evaluation of machine learning for ultrasonic NDE
- Research Program Outcomes

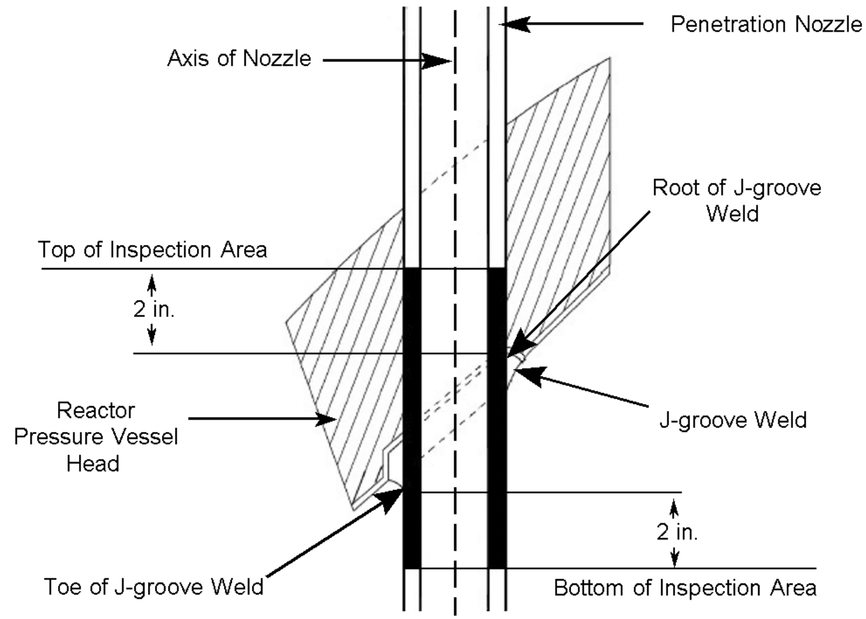
Acronyms

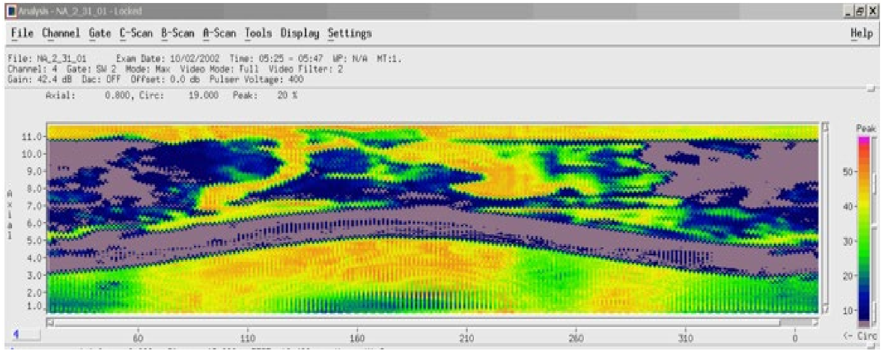
ADA – automated data analysis
ASME Code – American Society of Mechanical Engineers
Boiler and Pressure Vessel Code
CASS – cast austenitic stainless steel
CNN – convolutional neural network
CS- carbon steel
DMW – dissimilar metal weld
DNN – deep neural network
DR – detection rate
EPRI- Electric Power Research Institute
FPR – false positive rate
ISI – inservice inspection
ML – machine learning
NDE – nondestructive examination
ORNL – Oak Ridge National Laboratory
POD – probability of detection
PNNL – Pacific Northwest National Laboratory
ROC – Receiver Operating Curve
RVUH – reactor vessel upper head
TFC – thermal fatigue cracks
TPR – true positive rate
UT – ultrasonic testing (ultrasonics, ultrasonic examination, etc.)
UV – UltraVision
VP – VeriPhase
WSS – wrought stainless steel

Introduction and Background

Nondestructive Examination (NDE) in Nuclear Power Plants

- 10 CFR 50.55(a)(b) incorporates by reference the American Society of Mechanical Engineers Boiler and Pressure Vessel Code (ASME Code), Section III, *Rules for Construction of Nuclear Facility Components*, and Section XI, *Rules for Inservice Inspection of Nuclear Power Plant Components*
- NDE needed for timely detection of service-induced flaws
- Plant aging increases likelihood of service-induced flaws
- Accurate & Reliable NDE increasingly important due to industry trends to reduce:
  - Inspection time during outages
  - Radiation exposure
  - Number of examinations





Drivers for Automated Data Analysis (ADA)

- Section XI, Appendix VIII, Performance Demonstration for Ultrasonic Examination Systems, provides requirements for performance demonstration for ultrasonic examination procedures, equipment and personnel to detect and size flaws
- Industry projecting potential shortage in NDE technicians with proper skillsets to conduct NDE to meet future fleet needs (ML24026A087)
- Some UT inspections such as upper head exams yield large quantities of data that must be reviewed by multiple qualified inspectors during the outage period. (EPRI 3002023718)
  - High level of focus required for long periods of time
  - Human factors related to fatigue and momentary loss of focus can challenge reliability of results

ADA Is Coming

- Widely available, open-source ML tools have enabled the development and application of ML algorithms for many uses
- These tools are becoming more powerful and easier to use over time
- The nuclear industry is funding work to use these tools for automated data analysis algorithms to analyze NDE data

ONNX

GLUON

PyTorch

TensorFlow

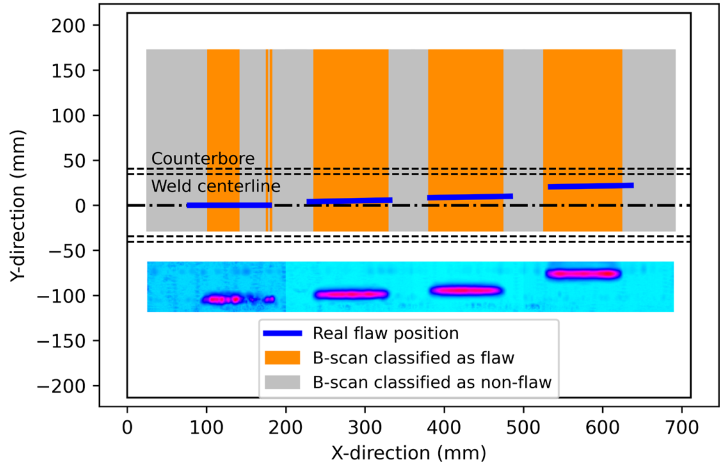
scikit learn

K

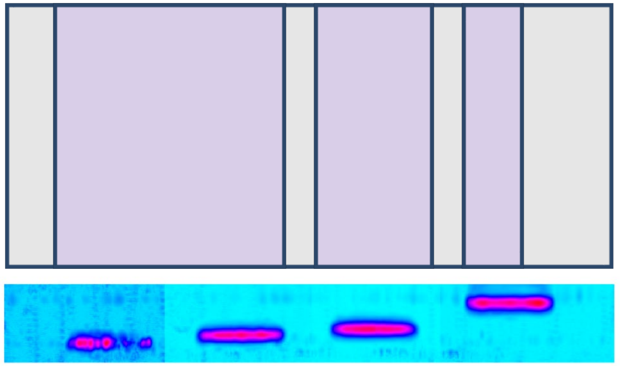
PyTensor

ADA/ML Use Cases for Ultrasonic NDE

- Near term
  - Analysis of encoded (recorded) data
  - Screening: Identify regions that are indication-free
  - Classification: Identify regions that contain flaws
  - Quality Control for NDE Examinations
- Longer term
  - Data compression
  - Generate NDE reports
  - Real Time data analysis of unencoded data
  - Synthetic data generation for training
  - …



Flaw Classification



Flaw Screening (Hypothetical Example)

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Two Ways of using ADA

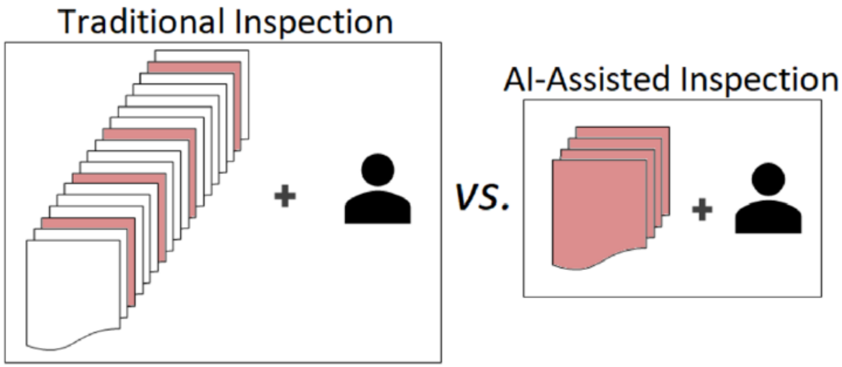
- ADA-Assisted Examination

  A fully-qualified inspector uses hints or highlighted areas to analyze the data, but the qualified individual makes the final calls

- Fully-Automated Examination

  The ADA algorithm makes the calls without human input

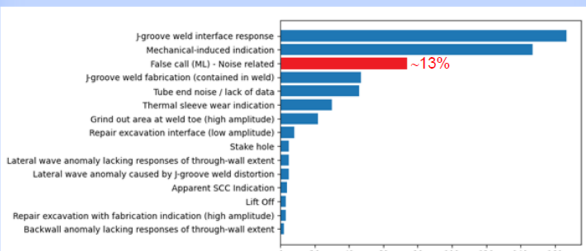
Automated Data Analysis -Assisted Procedures



Amount of Data Requiring Review

Pre-AI		Post-AI	
4.4	miles	463	feet
7.0	km	141	meters

All values are approximate



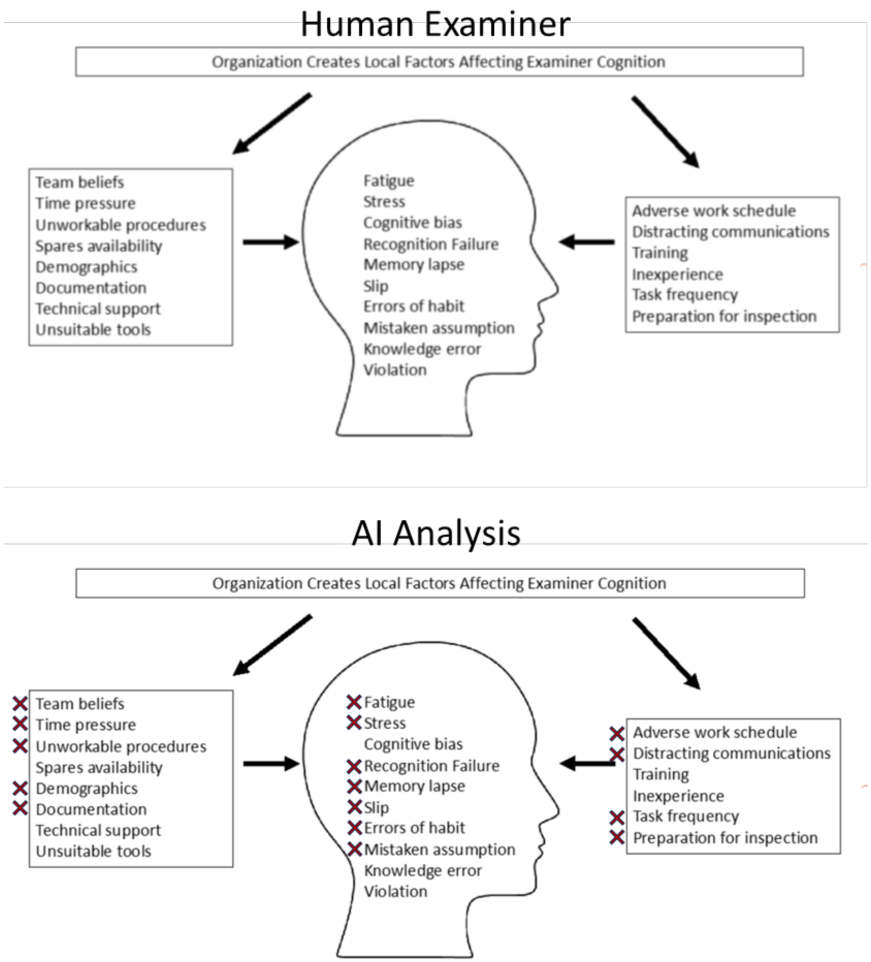
- One suggested approach by EPRI is for an ADA algorithm to flag areas with flaws, and the algorithm must find all flaws in the qualification set
- The algorithm can produce more false calls than allowed in the given supplement
- It will be up to the inspector to determine which of the areas flagged by the algorithm contain flaws, and ultimately the inspector is responsible for the results

Graphics from EPRI 3002023718

Automated Data Analysis – Possible Benefits

ADA has the potential to improve detection of flaws and improve the human factors of an examination.

- In-service flaws are rare in the nuclear industry. Computers can maintain vigilance in cases where humans struggle.
- Humans and computers make different types of mistakes, and a qualified analyst paired with an analysis run by ML gives the best of both worlds.
- Reduced dose to inspectors if ML used to support manual UT examinations.



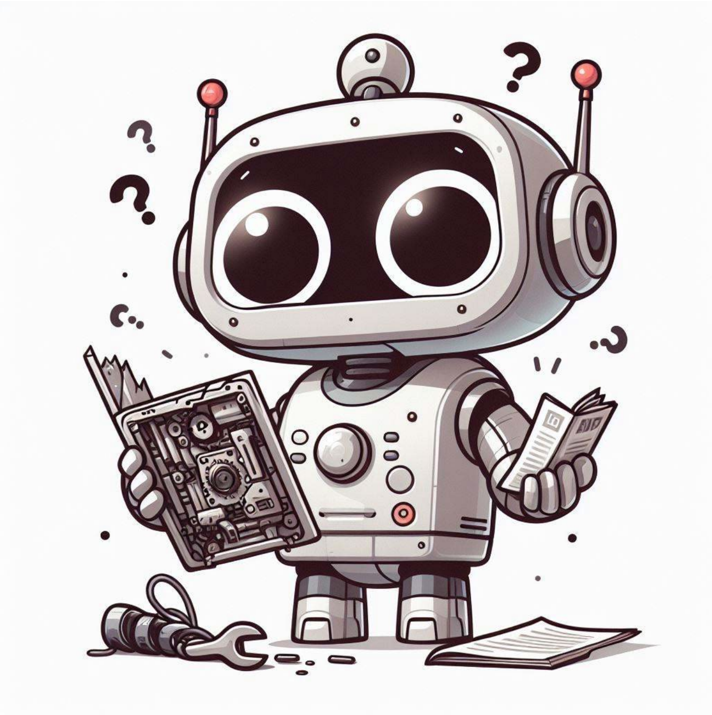
Graphic adapted from NUREG/CR-7295

Automated Data Analysis – Possible Hazards

- ADA has the potential to introduce common-cause failures of inspections across the fleet
- Licensees may not understand the capabilities and limitations of ADA, which could lead to improper use of ADA
- ADA assistance may allow people to pass Appendix VIII qualification testing without the skills to recognize unknown degradation in the field
- ML algorithms can be challenging to train and retrain, possibly making the ML algorithms unreliable
- ML algorithms require a new class of experts to support UT examinations

Automated Data Analysis – Expect the Unexpected

- As plants age and new reactors are designed, it is almost certain that new degradation mechanisms will emerge, and flaws will appear in unexpected places
- ADA methods can be very good at handling known problems but may not work on new forms of degradation



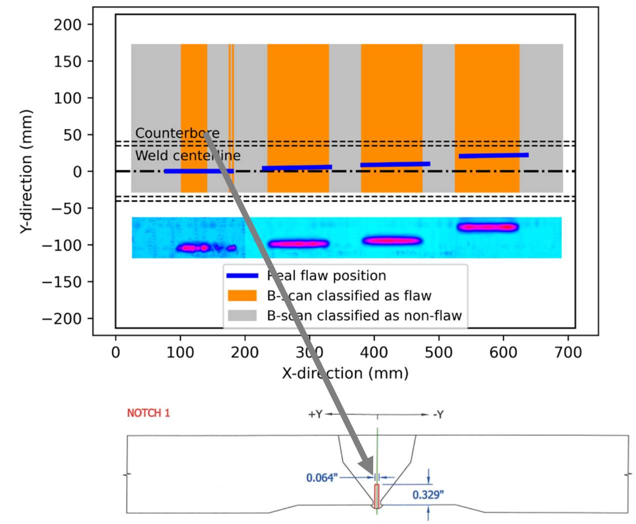
Research Program

Research Program on Automated Data Analysis

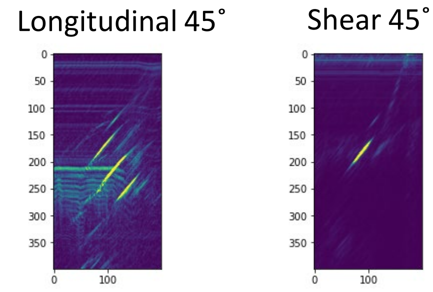
User Need Request for Evaluating the Reliability of Nondestructive Examinations, (NRR-2022-007), Task 4, Automated Data Analysis, requests that RES provide a technical basis describing current capabilities of machine learning and automated data analysis for nondestructive examination (NDE).

RES activity to address UNR request:

- Evaluating machine learning (ML) for Ultrasonic Examinations (UT) - Oak Ridge National Laboratory (ORNL)
- Evaluate commercially available automated data analysis platforms including rule-based and ML-based systems - Pacific Northwest National Laboratory (PNNL)



Flaw inside the weldment



Variation in Data (probe/mode)

Automated Data Analysis – Types of Algorithms

Rule-based

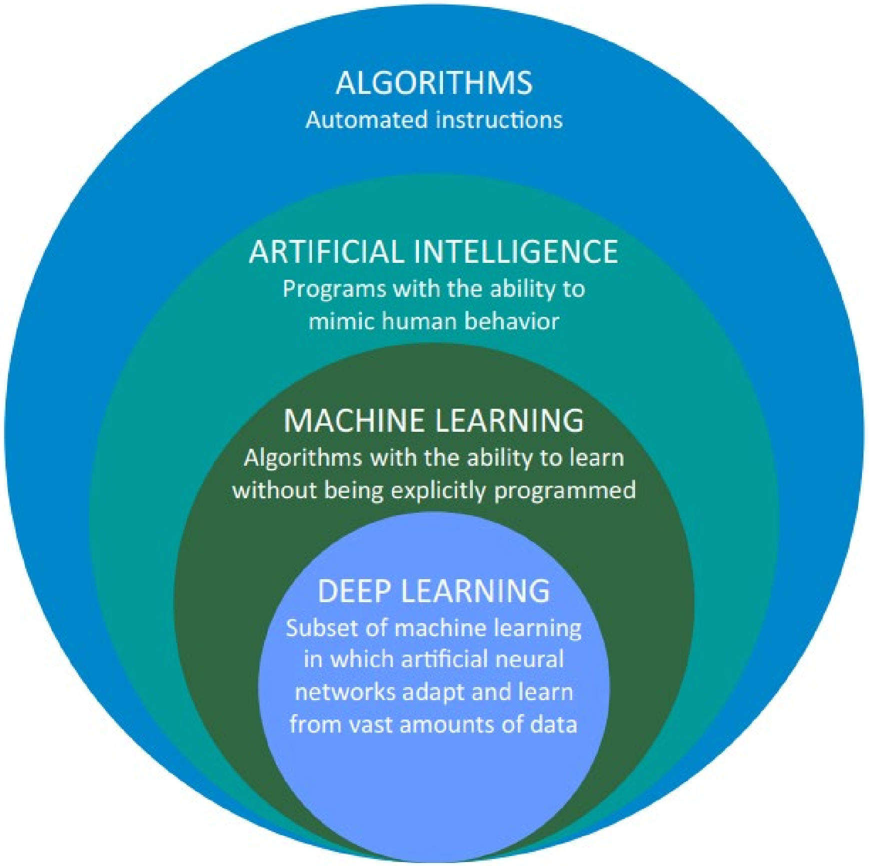
- Decisions made based off explicit rules
- Easy to determine why specific decisions are made

Learning-based

- Decisions based off training data
- Difficult to determine why specific decisions are made

Analysis

- Assisted – ADA provides analyst with flagged dataset
- Automated – No analyst



Vrana and Singh, 2021, https://doi.org/10.1007/s10921-020-00735-9

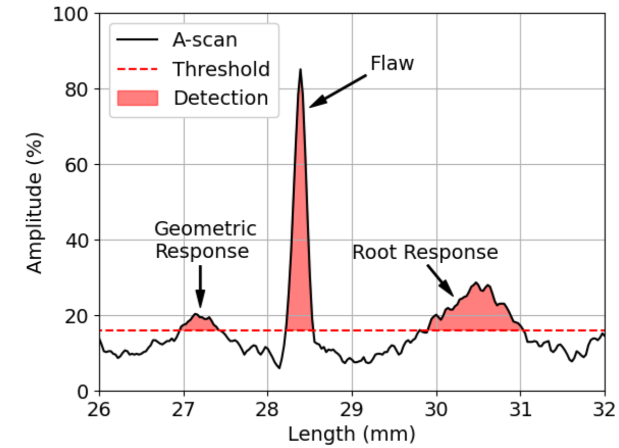
Evaluation of ADA for UT

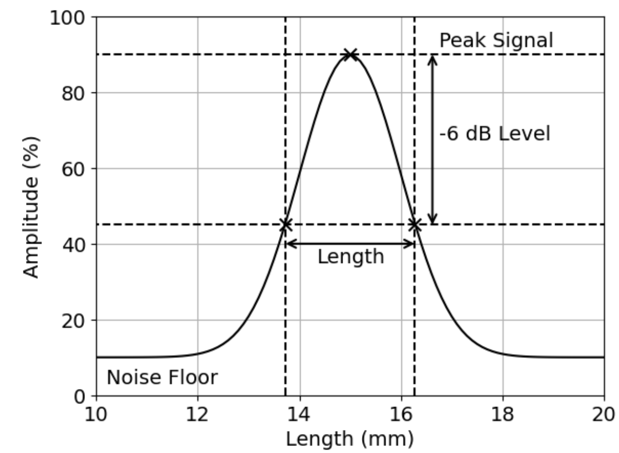
- Objectives
  - Assess current capabilities of ADA for improving NDE reliability
  - Provide technical basis to support regulatory decisions and Code actions related to ADA for NDE
- Expected outcomes
  - Identify capabilities and limitations of ADA for UT NDE applications
  - Identify factors influencing ADA performance and their impact on NDE reliability
  - Recommend verification and validation approaches and methods for qualifying ML (and ADA, as appropriate) for nuclear power NDE
  - Identify gaps in existing Codes and Standards relative to ADA for UT NDE

Assessment of Rule-Based ADA

Takeaways from Literature Review

- Almost all recent publications are dealing with learning-based analysis
- Rule-based ADA is usually used for flaw detection and signal processing
  - An amplitude threshold can be used to identify flaw signals above the noise floor
  - Signal processing can help improve signal to noise ratio
- Rule-based ADA can achieve high detection rates but also high false call rates
  - Not able to consistently distinguish between geometric and flaw responses

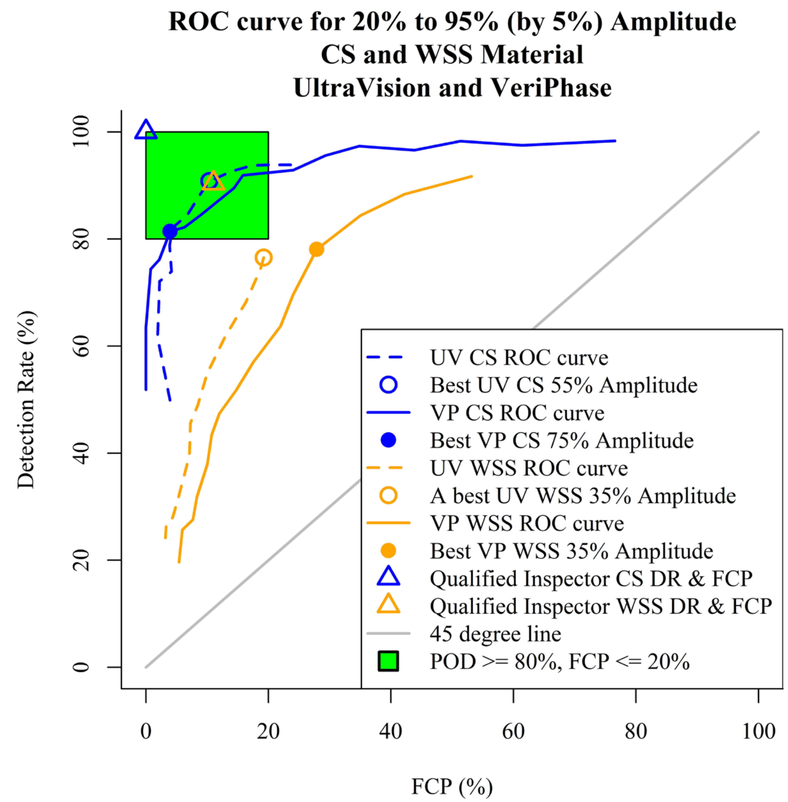




Assessment of Rule-Based ADA

Empirical Evaluation of Commercial ADA Systems

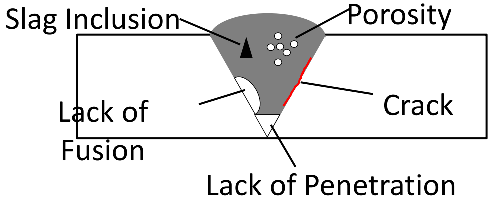
- Data analysis with two different commercial ADA software packages compared to analysis by qualified Level III UT analyst
- Statistical analysis of results using established methodologies



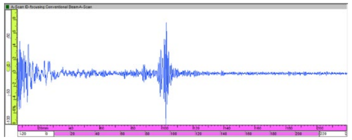
- Rule-based ADA is likely not fit for nuclear pipe inspections on its own
- Rule-based ADA could potentially be used alongside learning-based methods depending on the use-case

Assessment of Machine Learning (ML) Algorithms

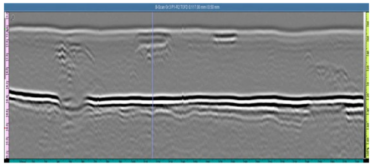
- Limited to ultrasonic NDE classification problems with data from weld inspections
  - Materials: Steel (*austenitic stainless steel, DMW, etc.*)
  - Flaw types: saw cuts, *EDM notches, thermal fatigue, stress corrosion cracking, weld fabrication flaws*
  - Inspection procedure assumed to be appropriate for weld inspections



https://www.zetec.com/blog/destructive-and-nondestructive-testing-of-welds-how-ndt-ensures-quality/

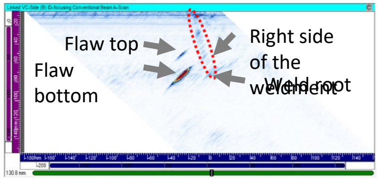


Example A-Scan



Example TOFD Scan

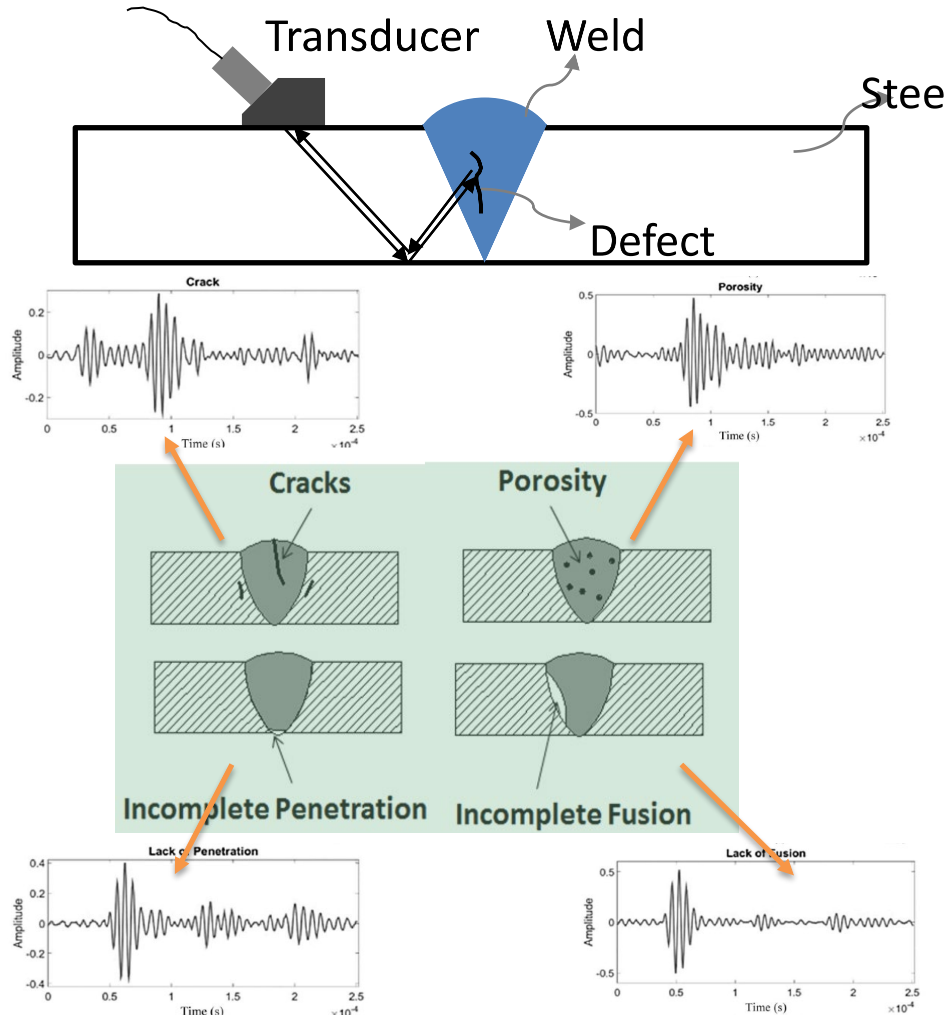
https://www.olympus-ims.com/en/applications/introduction-to-time-of-flight-diffraction-for-weld-inspection/



Example B-Scan

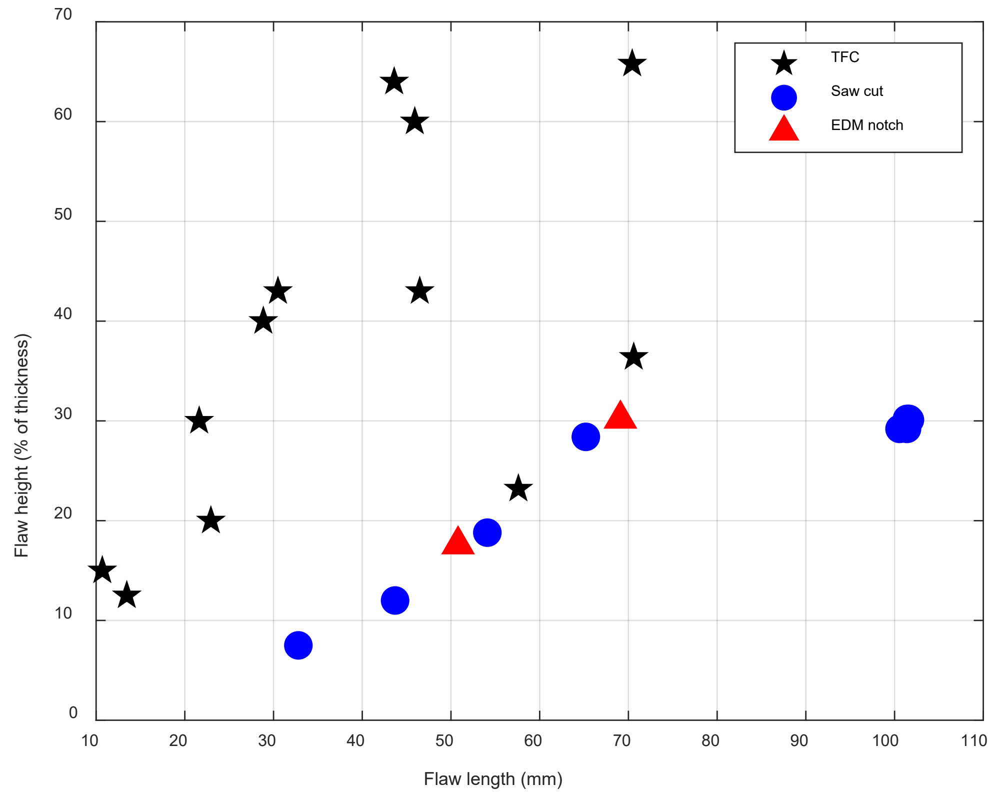
Empirical ML Research Objectives

- Determine capabilities and limitations of ML for NDE
- Identify factors influencing applicability to other inspections (CASS, DMW, RVUH, etc.)
- Assess effects of data augmentation, including using simulated data
- Establish methods to quantify confidence in ML results
- Assess capabilities for flaw size quantification from UT data



Generic Workflow for Assessment of ML for UT NDE

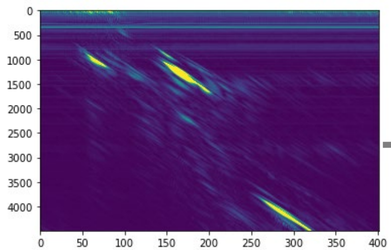
1. Collect ultrasonic NDE data from a variety of materials with multiple probe designs, frequencies and wave modes
2. Pre-process the data to remove noise and outliers
3. Train a machine learning algorithm on the preprocessed data
4. Use the trained algorithm to analyze new ultrasonic data
5. Assess the results using multiple metrics



Flaw size distribution for four stainless steel and two DMW specimens.

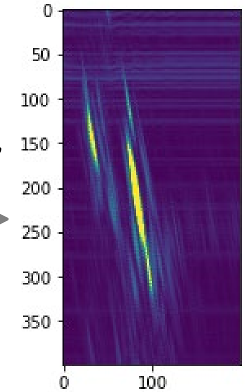
Overview of Empirical Assessment

Original B-scan



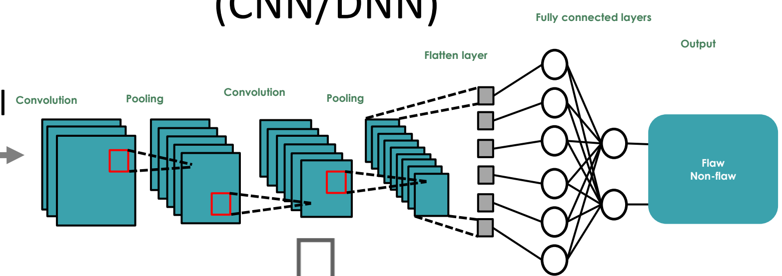
Preprocessing
(Crop, normalization, downsampling, flip)

B-scan input



Input to ML model

ML modeling
(CNN/DNN)

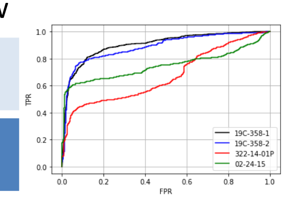


Accuracy=0.88
TPR=0.83
FPR=0.05

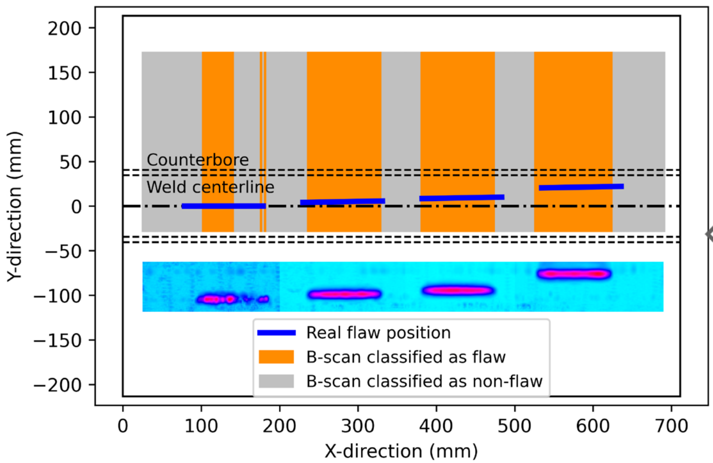
Accuracy, true/false positive rate
(TPR, FPR)

	Flaw	Non-flaw
Predicted as Flaw	339	14
Predicted as non-flaw	67	274

Confusion matrix



ROC curves



Visualize Results

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Examples of Results

Training with →

A	**Plate**	**4 saw cuts**
B	Plate	4 saw cuts
C	Pipe	3 TFC
D	Pipe	4 saw cuts 3 TFC

Specimen B

	Flaw	**Non-Flaw**
Predicted as **Flaw**	**339**	**14**
Predicted as **Non-flaw**	**67**	**274**

Accuracy=0.88
True positive rate (TPR)=0.83
False positive rate (FPR)=0.05

Specimen C

	Flaw	Non-Flaw
Predicted as Flaw	**40**	**27**
Predicted as Non-flaw	**88**	**325**

Accuracy=0.76
TPR=0.31
FPR=0.08

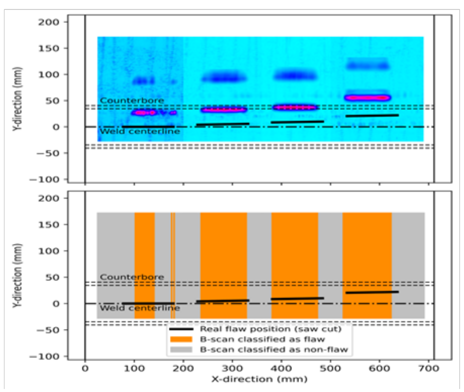
Specimen D

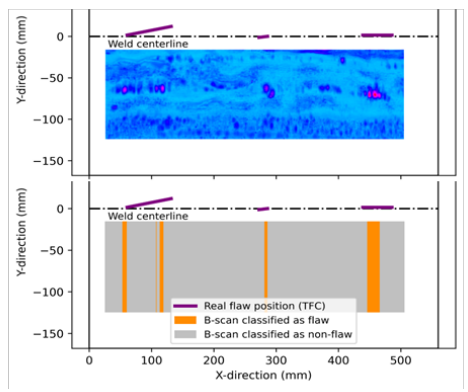
	Flaw	Non-Flaw
Predicted as Flaw	**243**	**58**
Predicted as Non-flaw	**29**	**352**

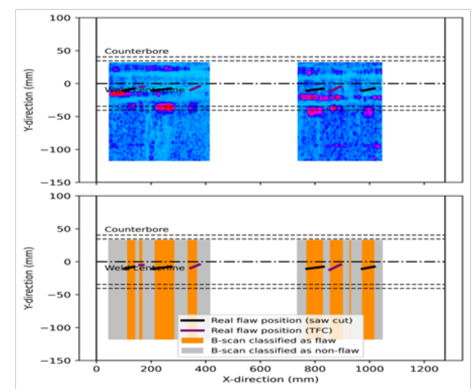
Accuracy=0.88
TPR=0.89
FPR=0.14

UV (top view)

Scan area







Low true positive rate on flaws close to weld centerline and on smaller TFC flaws

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Transfer Learning Example

Test results using the retrained model

Specimen A (4 saw cuts)

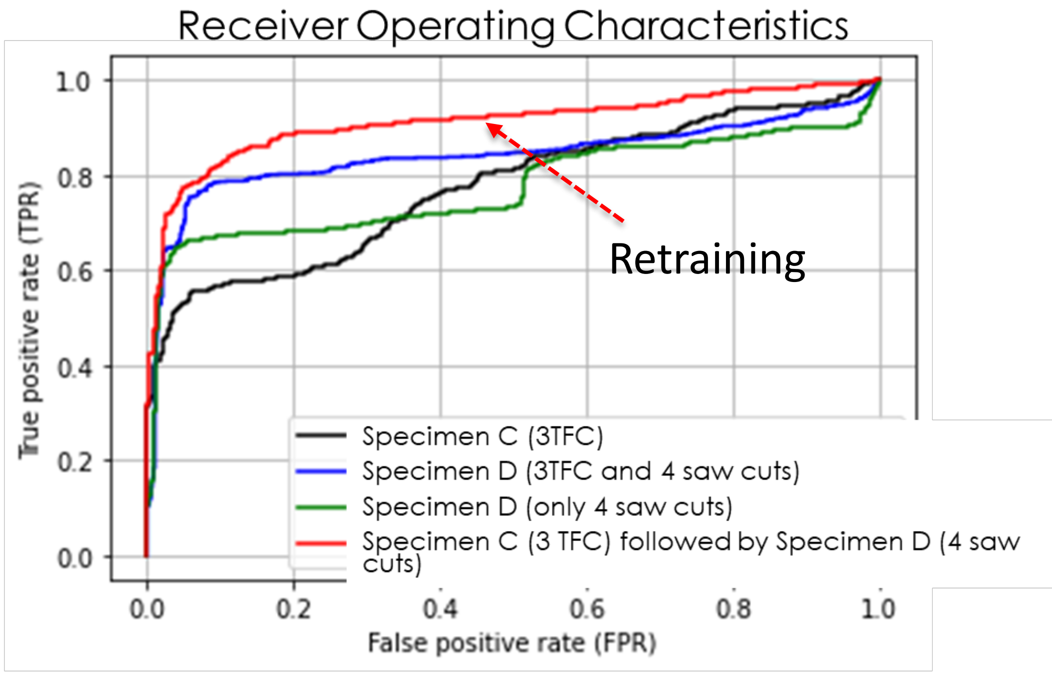
	Flaw	Non-Flaw
Predicted as Flaw	**362**	**26**
Predicted as Non-flaw	**44**	**262**

Accuracy=0.90
TPR=0.88
FPR=0.09

Specimen B (4 saw cuts)

	Flaw	Non-Flaw
Predicted as Flaw	**338**	**65**
Predicted as Non-flaw	**68**	**223**

Accuracy=0.81
TPR=0.83
FPR=0.22



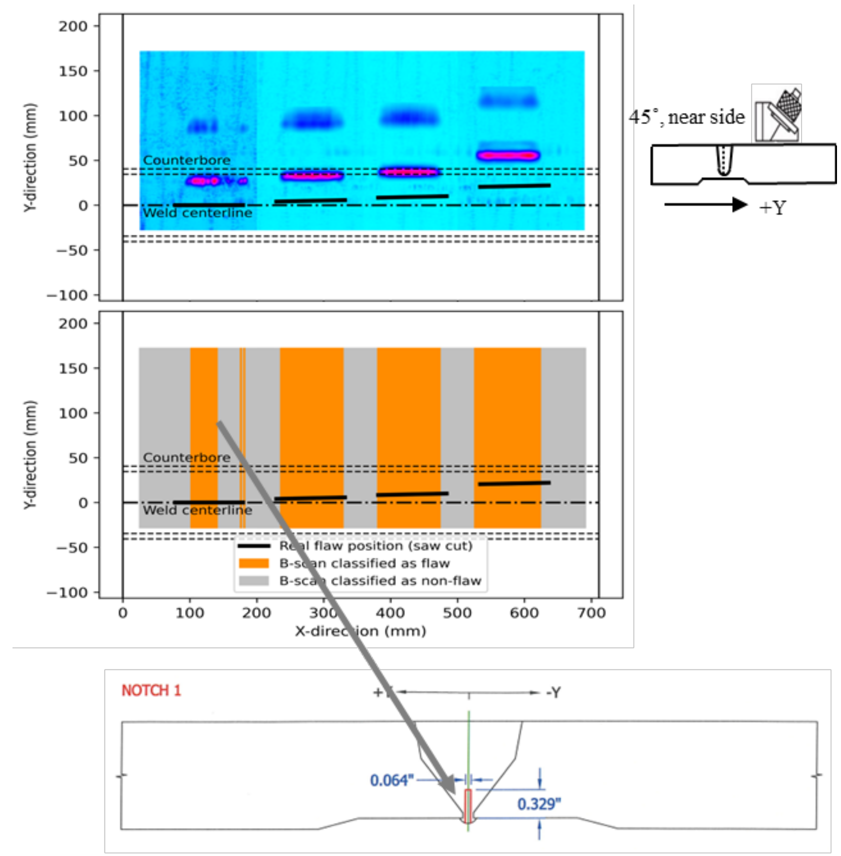
ROC curves on specimens A and B from different models

Retraining and incorporating transfer learning methods may help to improve the performance when the model encounters new data.

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Findings to Date: ML

- Capable of high TP, low FP and FN
- May be able to learn key signatures using data from simple flaws (e.g. saw cuts) and generalize well to other flaw types (e.g., TFC)
  - Generalization capability may vary with flaw size and location
- Transfer learning techniques may be useful for improving accuracy with new data sets
- Model type (for instance, NN vs DNN) may not *significantly* change results



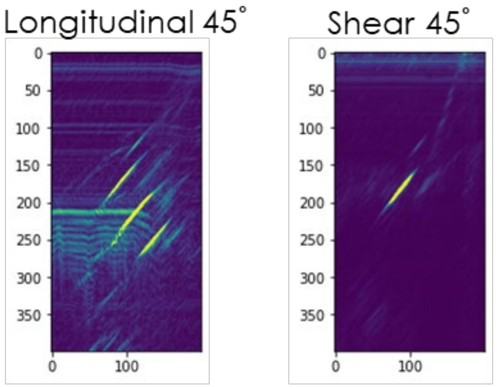
Flaw inside the weldment (TPR=0.54)

ML , *if used with care*, can be used for NDE data classification

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Findings to Date: Data

- Training data should be representative of the types of data expected during testing
  - Expanded training data sets may allow ML to accommodate nominal weld geometrical variances and associated noise

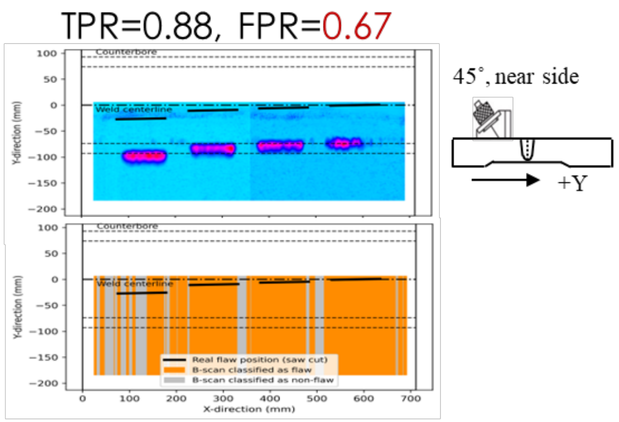


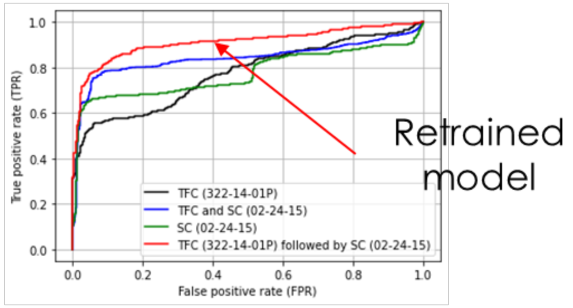
Variation in Data (probe/mode)

- High accuracy possible if test data is “in distribution” relative to training data
  - Consistency across training and test data sources important for high classification accuracy

Findings to Date: Metrics

- Desired performance thresholds likely dependent on use case
- Commonly used metrics: TPR, FP, FN
  - Low FP and FN rates, high TPR desirable
  - Zero FN, low FP, high (100%) TPR for screening?
- Other useful measures
  - Receiver operating characteristic (ROC) curves – TPR vs FPR
  - ML training curves – can indicate overfitting and potential poor classification accuracy if deployed



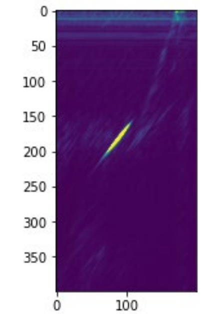


U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

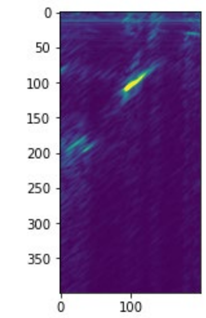
Findings to Date: Best Practices

- Consistency in preprocessing procedures (crop, normalization, down-sampling, etc.)
- Review and correct, if necessary, output labels
- Tuning and selecting parameters that control the learning method
- Retraining a trained network with additional data to improve performance and tune ML to site-specific data

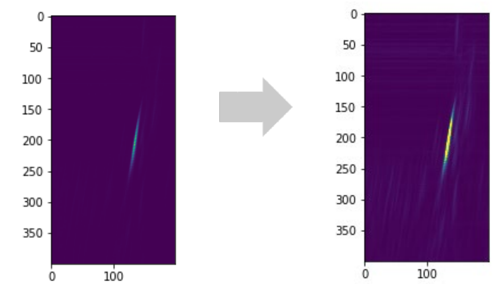
B-scan of saw cut



B-scan of TFC

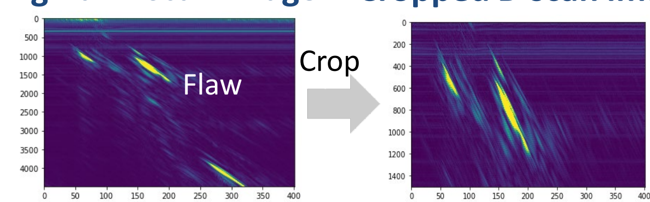


Normalized to 40 dB



Original B-scan Image

Cropped B-scan image



Status of RES Program – Assessment of Commercially-available Algorithms/Systems

- Technical Letter Report entitled "Evaluation of Commercial Rule-Based Assisted Data Analysis" in the RES/NRR review cycle
- Confirmatory analysis of the commercial ML system being tested by industry in field trials has recently begun
  - Focus on upper head examinations
  - Mockups being designed and fabricated
  - Assessment will include:
    - Pre-trained algorithm tested with vendor collected UT data on NRC-owned mockups
    - Training and testing with PNNL/ORNL data with comparison of results to ORNL ML algorithm results

Status of RES Program – ML for UT NDE Ongoing Research

- Impact of ML on POD, and comparison of ML results with manual analysis performed by a qualified analyst (including comparison of ML performance against Appendix VIII requirements)
- AI-Assisted vs Fully-Automated analysis: Detection and sizing of degradation that the ML system has not been trained on, validation/qualification requirements, and essential variables
- Qualification of ML
  - Training, test, validation data requirements, and benchmark data sets
  - Acceptable performance thresholds and requalification processes
- Methods for establishing confidence in ML results
  - Verification and validation of data and methods
  - Uncertainty quantification, ML interpretability, and related criteria (if any) for qualification

Status of RES Program – ML for UT NDE

- Technical letter report entitled "An Assessment of Machine Learning Applied to Ultrasonic Nondestructive Evaluation" (ORNL/SPR-2023/3245) published February 2024 (ML24046A150)

- Other publications
  - H. Sun, R. Jacob, and P. Ramuhalli, "Classification of Ultrasonic B-Scan Images from Welding Defects Using a Convolutional Neural Network," *Proc. 13th NPIC&HMIT 2023,* Pages 272 - 281 . ISBN 978-0-89448-791-0 (ML23241A961)
  - H. Sun, P. Ramuhalli, and R. Jacob, "Machine Learning for Ultrasonic Nondestructive Examination of Welding Defects: A Systematic Review," *Ultrasonics,* Vol. 127 Issue 1, Jan 2023, Pages 106854 (ML22284A071)

Research Program Outcome – providing the technical basis to answer...

Potential Qualification Pathways for ADA (including ML)

ADA for classification (flaw detection)

- Can adopt approach similar to existing Section XI, Appendix VIII for performance demonstration
- Assumed standard for performance:
  - Greater than or equal to current practice (i.e. human performance)
  - Could adopt similar acceptance criteria for performance demonstration

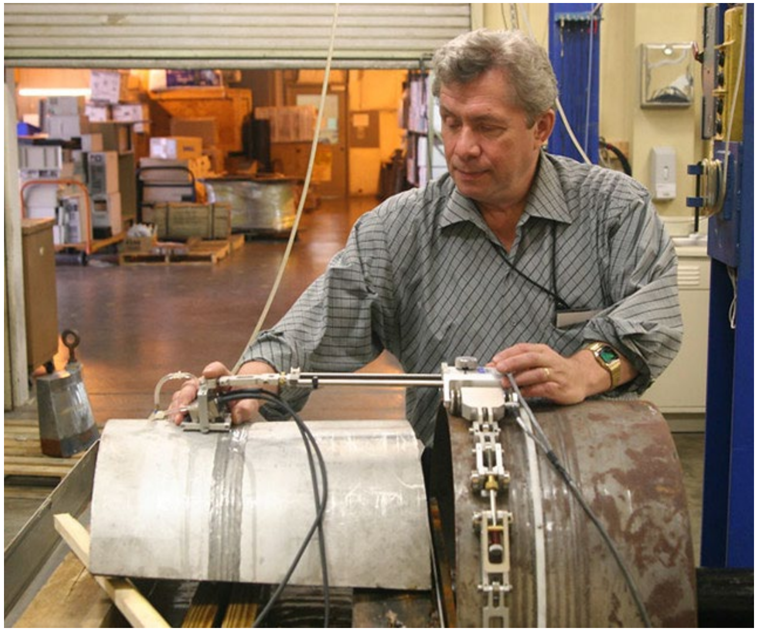
ADA for screening (excluding unflawed regions from evaluation)

- Can adopt approach similar to existing Section XI, Appendix VIII for performance demonstration
- Biased toward calling "detections"
  - Goal is to have **no** "misses"
  - Tolerance for high false call rate
  - Qualified UT analyst responsible for all calls
- Acceptance criteria should reflect the bias toward detection
- Do training/qualification specimens need to incorporate non-flaw features intended to generate a "detection" response with the algorithm?

If ML-based ADA has the potential to be better than current practice, then should ADA be held to a higher performance standard?

Initial Qualification Requirements for ADA-Assisted Examinations

- A UT procedure that uses ADA-assistance can currently be qualified using Appendix VIII as the user of the procedure is a UT Level II
- How should the qualification requirements specified in Section XI, Appendix VIII be updated?
  - Currently only covers encoded data
  - There are many complexities associated with training ML algorithms not captured in current rules



U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Implications Related to Retraining ADA Algorithms

- If an ML algorithm is retrained, the algorithm has been altered and is a change of an essential variable in the procedure
- In ASME Code Section XI Appendix VIII, a procedure must be requalified via a successful personnel qualification if an essential variable is changed
- The NRC understands the potential benefits of changing the ASME Code to allow for field-friendly implantation of ML (e.g. requalifying a retrained ML algorithm on-site)

Paths to the Future for ADA

Near Future on Current Trajectory

Skilled Inspectors + Engaged AI Experts + Optimized Procedures = Best Outcome

Current Inspections Improved

New procedures can be developed

New degradation can be found

New inspectors become overly-dependent on AI tools

AI experts move on to new tasks

Lack of AI experts prevents effective development and retraining of algorithms

Unskilled Inspectors + No AI Experts + Unoptimized Procedures = Worst Outcome

Current Inspections may be OK

New degradation may not be found

New Procedures may be challenging

Care must be taken to prevent this outcome.

Avoiding Future Problems

- Industry needs to build the infrastructure to allow for the effective use of ADA
- Create rules for requalifying an algorithm after modification that does not require a person to pass a personnel test
  - e.g. Finds all flaws in qualification data without too many additional false calls
- Requirements for personnel to use ADA-assisted procedures to assure that they have appropriate skills
  - e.g. Pass an Appendix VIII tests for the same Supplement without ADA assistance