Findings of a Literature Survey on Machine Learning for Nondestructive Examination

P. Ramuhalli, H. Sun, D. Womble, R. Jacob*

Oak Ridge National Laboratory

* Pacific Northwest National Laboratory

Second Data Science and AI Regulatory Applications Workshop

Aug 18, 2021

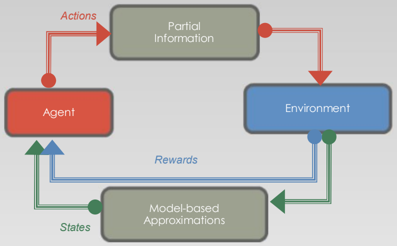
Sponsored by US NRC / Office of Research under IAA 31310020F0038.
Carol Nove, NRC COR

ORNL Strategic Directions in AI/ML

Data

- Facilities operation and control
- Experimental design
- Data curation and validation
- Compressed sensing

Learning

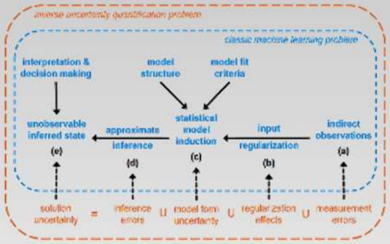


- Physics informed
- Accelerating learning
- Stability and robustness
- Foundations of ML formulations - RL, GANs, GNNs, BNNs
- Dimension reduction and encoding

Scalability

- Algorithms, complexity and convergence
- Levels of parallelization
- Mixed precision arithmetic
- Communication
- Implementations on accelerated-node hardware

Assurance



- Uncertainty quantification
- Robustness
- Explainability and interpretability
- Validation and verification
- Causal inference and hypothesis generation

Workflow

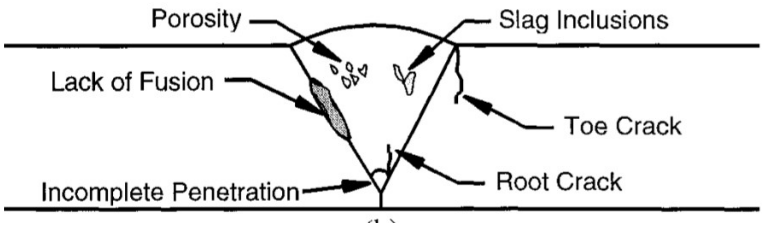
- Edge AI
- Compression
- Online learning
- Federated learning
- Infrastructure
- Augmented intelligence and Human-Computer Interface

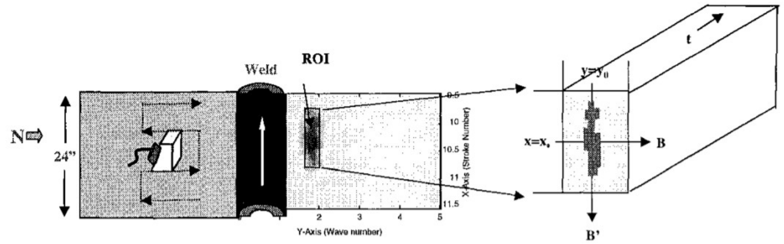
Outline

- Overview
  - Nondestructive examination (NDE)
  - Artificial intelligence (AI)/machine learning (ML)
- Machine learning for nondestructive examination
  - Background
  - Objectives
- Key findings from literature assessment
- Summary and next steps

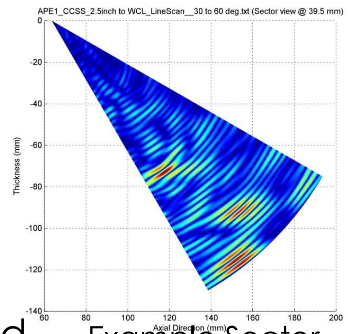
Nondestructive Examination (NDE) in Nuclear Power

- Detect surface or internal anomalies that could compromise the ability of a component to perform its function
  - Examination methods generally classified as volumetric, surface and visual
- Inservice inspection (ISI) of nuclear power plant components required by10CFR50.55a which incorporates by reference Sections III and XI of the ASME Boiler and Pressure Vessel Code
- Analysis of NDE examination data typically performed manually by qualified inspectors
- Increased interest in machine learning (ML) for flaw detection in ASME Code-required inspections
  - Anticipated cost savings, time savings, and expected future shortage of qualified inspectors
  - Potential for future Code activities in application of ML, and licensee submittals

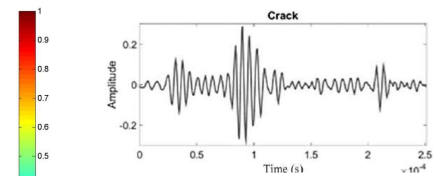




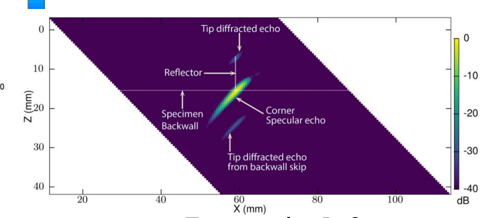
Weld Inspection Example (From J. Kim et al, QNDE 2001)



Example Sector Scan



Example A-Scan



Example B-Scan (From PNNL-26336)

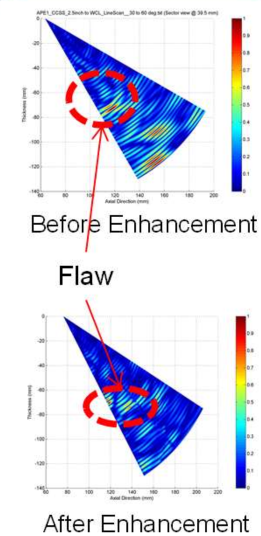
What is the impact of ML on NDE reliability?

Oak Ridge National Laboratory

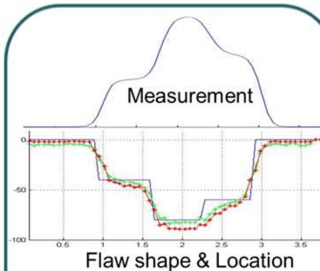
Machine Learning for NDE

- Objectives
  - Assess current capabilities of ML and automated data analysis for improving NDE reliability
  - Provide technical basis to support regulatory decisions regarding reviews of relief requests and Code actions that implement automated data analysis for NDE of nuclear power plant components
- Expected outcomes
  - Identify capabilities and limitations of ML for ultrasonic NDE applications
  - Identify factors influencing ML performance and their impact on NDE reliability
  - Recommend verification and validation (V&V) approaches and methods for qualifying ML for nuclear power NDE
  - Identify gaps in existing codes and standards relative to ML for ultrasonic NDE

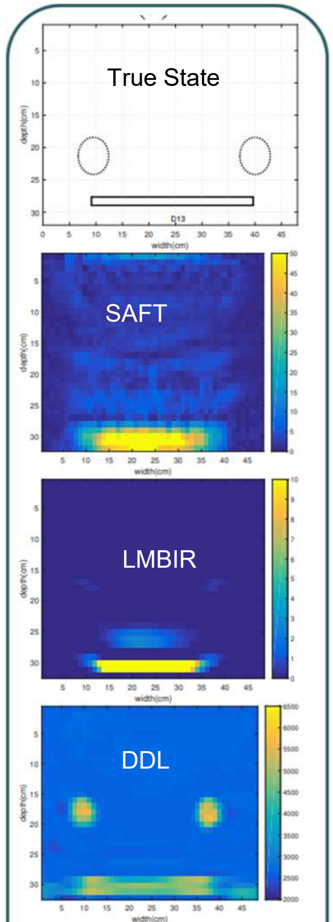
Examples



Flaw signal enhancement



Flaw Reconstruction*



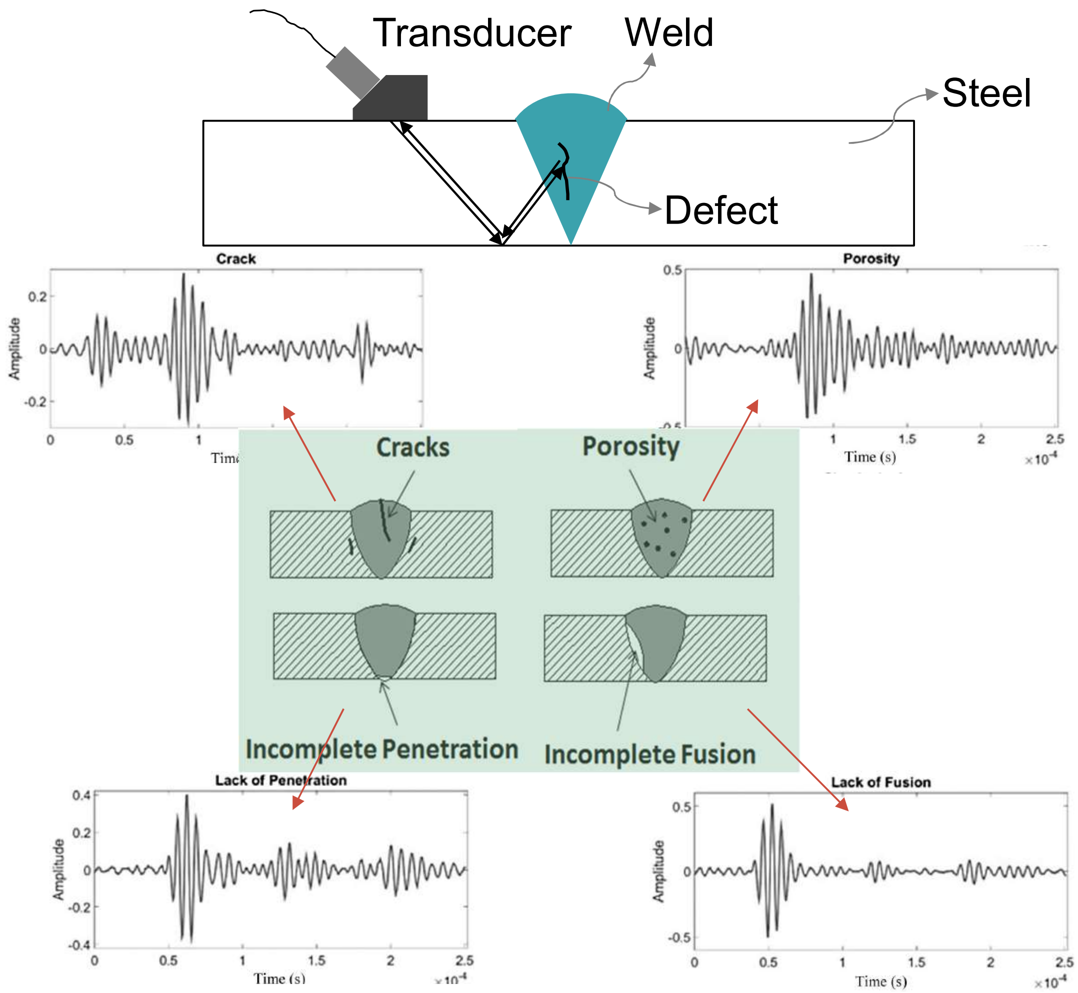
Flaw Reconstruction**

*T. Khan 2011 (PhD Dissertation, Michigan State U.)

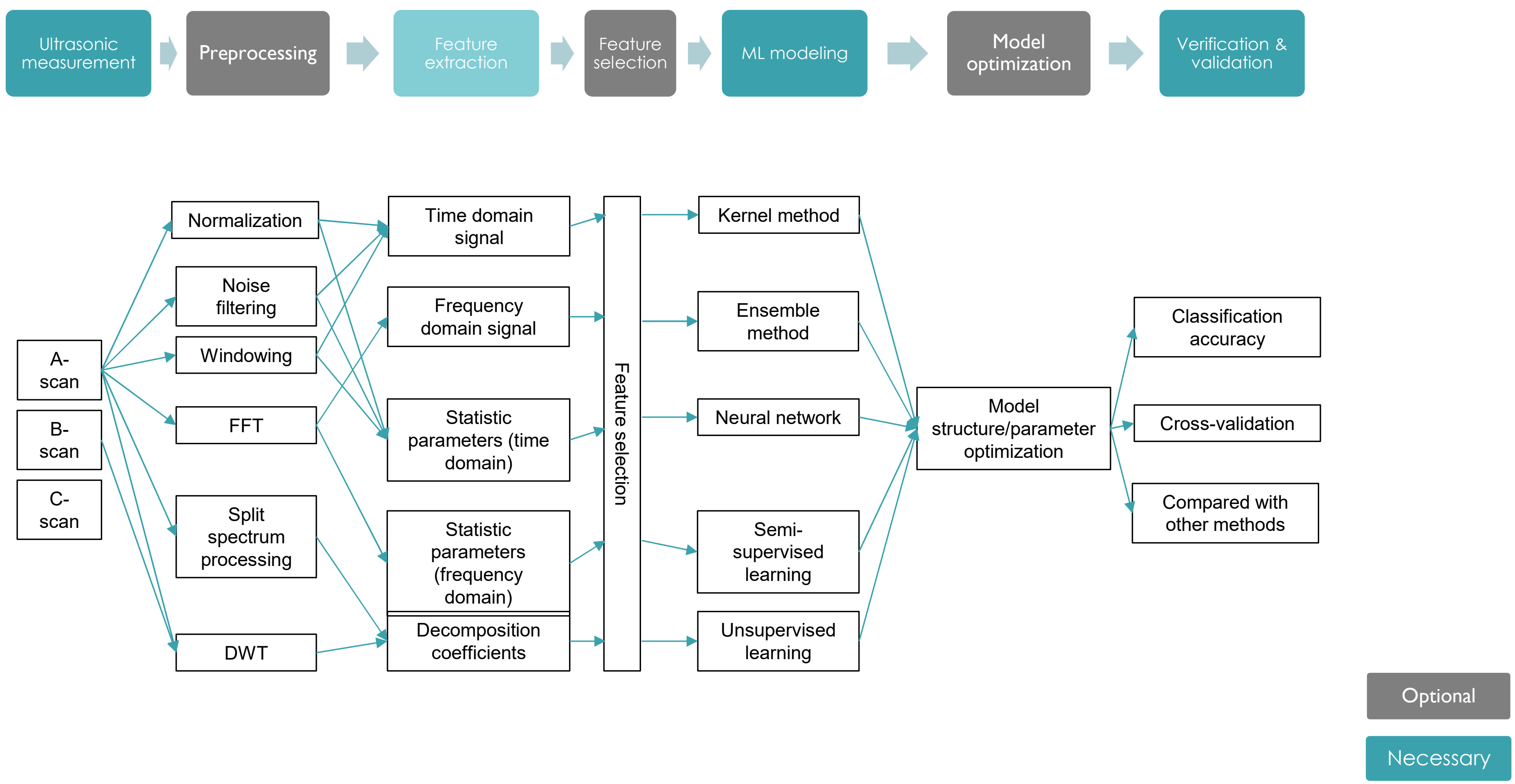
**H. Almansouri, S. V. Venkatakrishnan, G. T. Buzzard, C. A. Bouman and H. Santos-Villalobos, 2018 IEEE GlobalSIP; 10.1109/GlobalSIP.2018.8646704.

Focus: Ultrasonic NDE, Data-driven Learning Algorithms

- Limited to NDE classification problems with data from weld inspections
  - Materials: Steel (<u>*carbon, austenitic*</u>, cast,…), nickel alloys
  - Flaw types: *thermal fatigue, stress corrosion cracking, weld fabrication flaws*
  - Inspection setup assumed to be appropriate for weld inspections
- Approach: <u>**Literature review**</u> followed by empirical studies
  - Literature set identified is large but not exhaustive

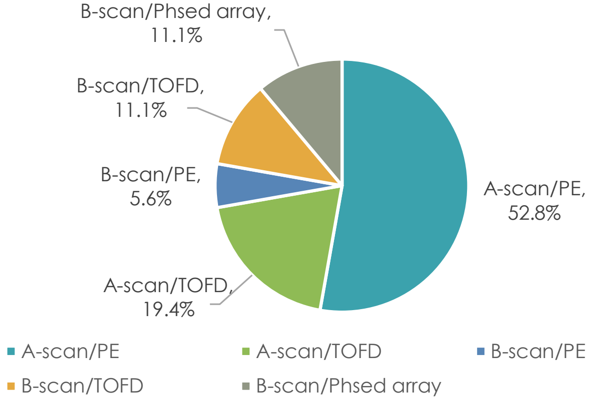


Data Flow in ML for Ultrasonic NDE

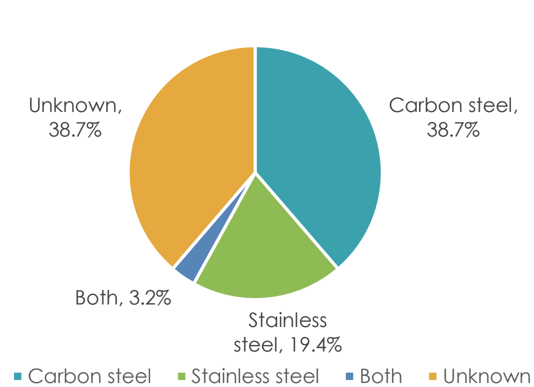


Summary of Literature Data

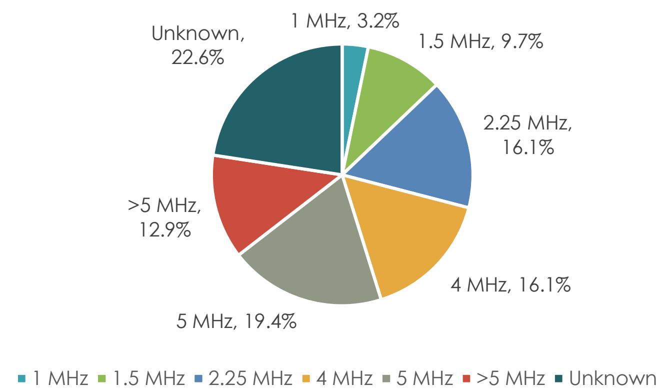
Ultrasonic measurement type



Specimen materials

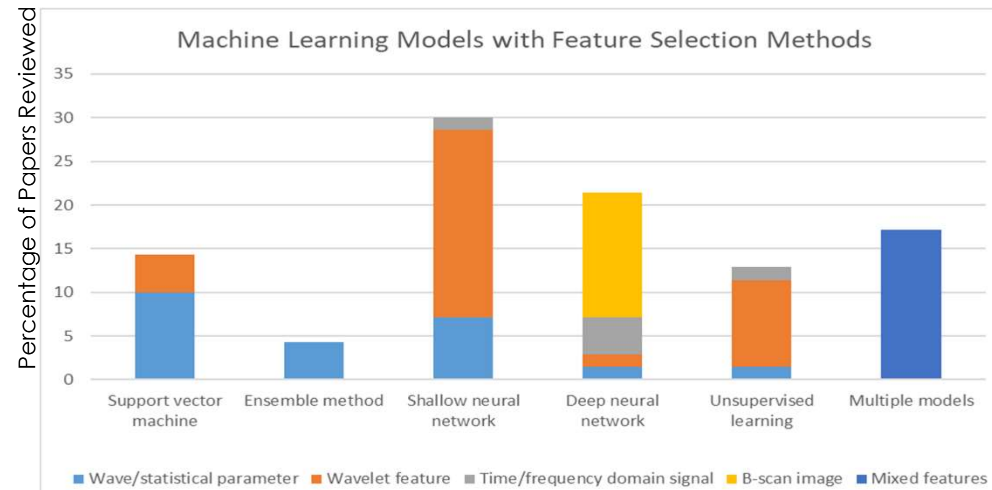


Ultrasonic NDE: Frequencies



Total instances	Physical specimens	Number of crack flaws	Number of non-crack flaws	RT verification?	Ref.
50	Steel specimen	15	35	N/A	[19]
273	Simulated flaws	73	100	N/A	[5]
240	10 steel specimens	N/A	N/A	Yes	[28]
282	1 steel specimen	N/A	N/A	N/A	[29]
100	100 steel specimens	0	100	N/A	[30]
61	Bearing steel samples	N/A	N/A	N/A	[15]
438	438 specimens	0	217	N/A	[14]
239	Steel specimen	N/A	N/A	N/A	[43]
135	135 specimens	45	90	Yes	[7]
246	EPRI database	N/A	N/A	N/A	[16]
293	EPRI database	N/A	N/A	N/A	[16]
90	90 steel specimens	25	44	N/A	[31]
120	6 aluminum specimens	N/A	N/A	N/A	[6]
240	12 steel specimens	N/A	N/A	N/A	[8]
200	12 steel specimens	N/A	N/A	N/A	[8]

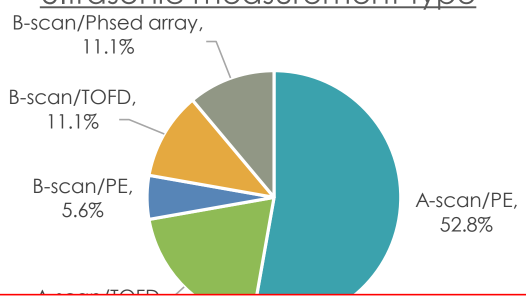
Examples of Data Distribution



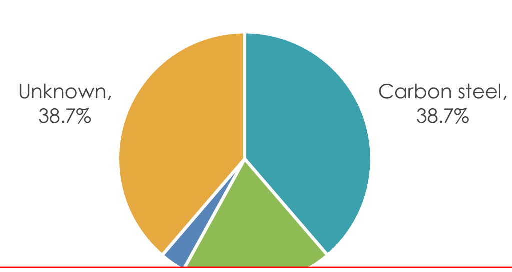
ML Models and Feature Selection

Summary of Literature Data

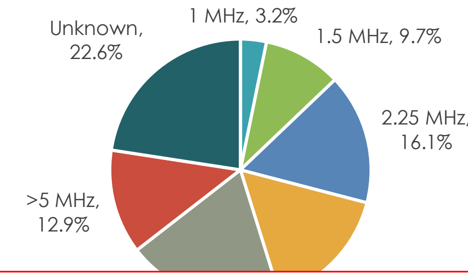
Ultrasonic measurement type



Specimen materials



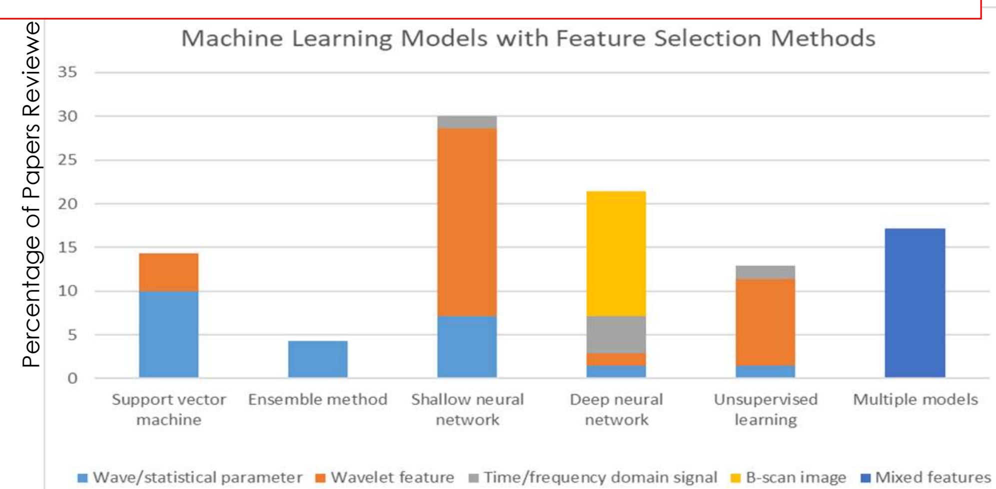
Ultrasonic NDE: Frequencies



Lack of common data sets and diversity in methods/data sets challenge direct comparisons, though general insights into capabilities possible.

Total instances	Physical specimens	Number of crack flaws	Number of non-crack flaws	RT verification?	Ref.
50	Steel specimen	15	35	N/A	[19]
273	Simulated flaws	73	100	N/A	[5]
240	10 steel specimens	N/A	N/A	Yes	[28]
282	1 steel specimen	N/A	N/A	N/A	[29]
100	100 steel specimens	0	100	N/A	[30]
61	Bearing steel samples	N/A	N/A	N/A	[15]
438	438 specimens	0	217	N/A	[14]
239	Steel specimen	N/A	N/A	N/A	[43]
135	135 specimens	45	90	Yes	[7]
246	EPRI database	N/A	N/A	N/A	[16]
293	EPRI database	N/A	N/A	N/A	[16]
90	90 steel specimens	25	44	N/A	[31]
120	6 aluminum specimens	N/A	N/A	N/A	[6]
240	12 steel specimens	N/A	N/A	N/A	[8]
200	12 steel specimens	N/A	N/A	N/A	[8]

Examples of Data Distribution

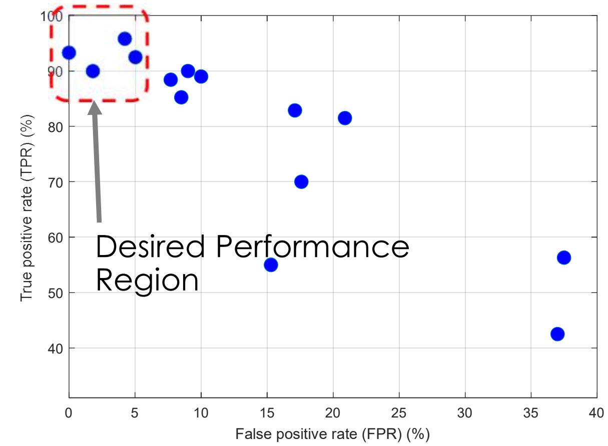


ML Models and Feature Selection

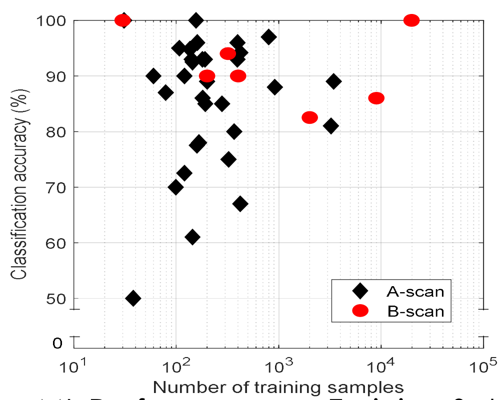
OAK RIDGE National Laboratory

Examples of Reported ML Performance in the Literature

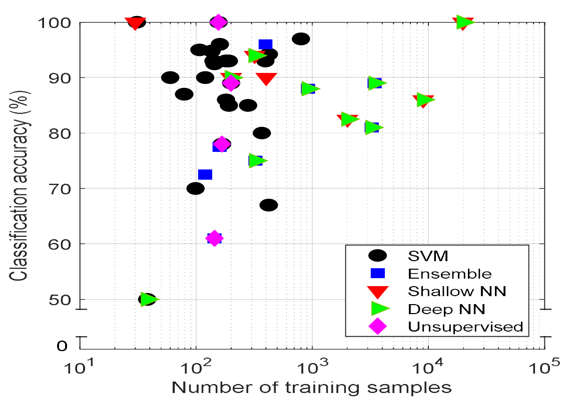
What factors influence the performance of machine learning (ML) and automated data analysis techniques when applied to NDE data?



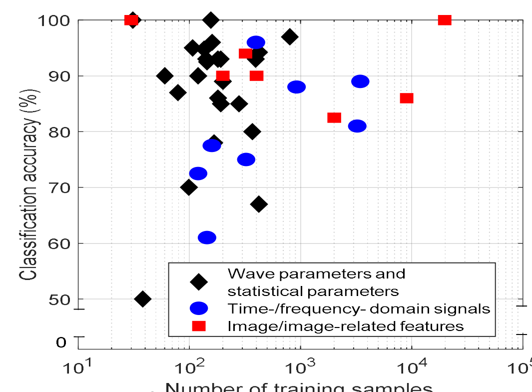
True Positive Rate (TPR) vs False Positive Rate (FPR)



ML Performance vs Training Set Size, Sorted by Type of Data



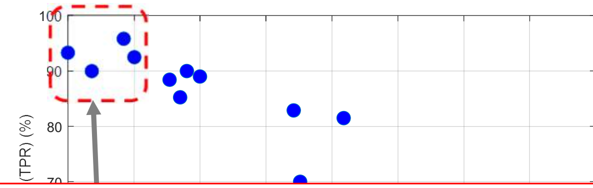
ML Performance vs Training Set Size, Sorted by ML Method



ML Performance vs Training Set Size, Sorted by Feature Type

Examples of Reported ML Performance in the Literature

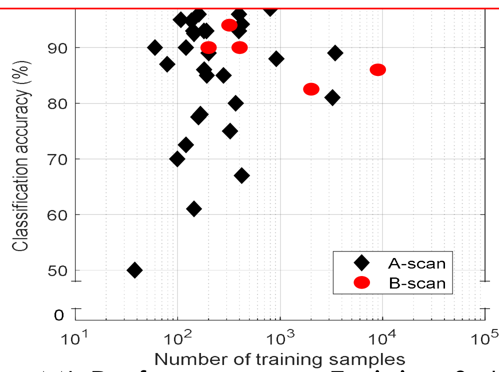
What factors influence the performance of machine learning (ML) and automated data analysis techniques when applied to NDE data?



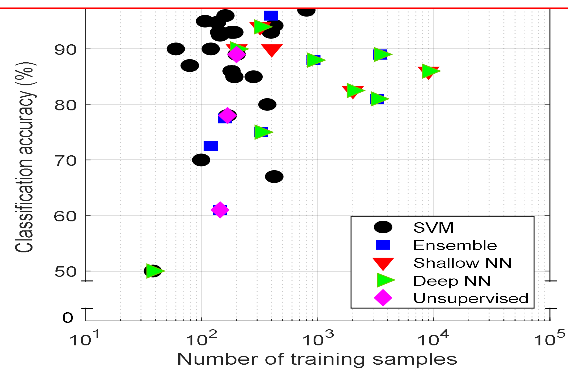
High classification accuracy (high true positive rate and low false positive/negative rate) is possible with ML applied to ultrasonic NDE data

Most ML methods are likely to be capable of good classification performance, with performance depending on the data used for model training and model parameter tuning

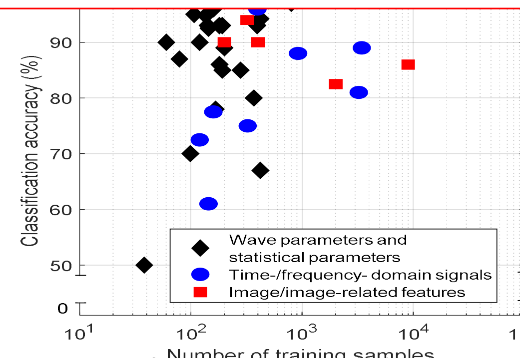
There may be a need for common data sets to compare across methods/solution providers



ML Performance vs Training Set Size, Sorted by Type of Data



ML Performance vs Training Set Size, Sorted by ML Method

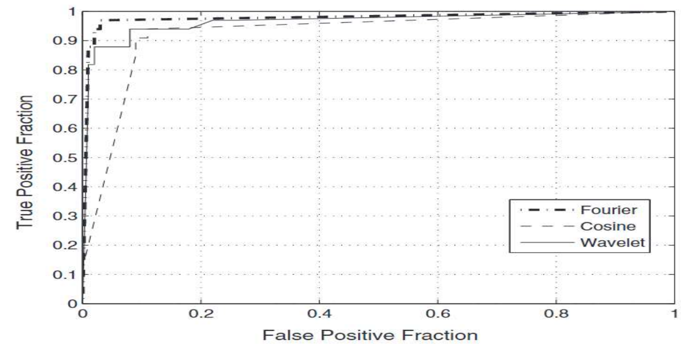


ML Performance vs Training Set Size, Sorted by Feature Type

ML and Ultrasonic NDE Reliability

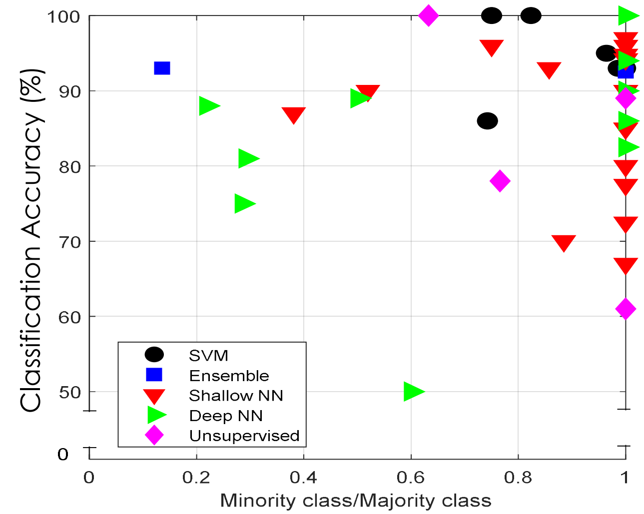
- Limited information in literature on:
  - Sensitivity of classification performance to various factors
  - Demonstrating confidence in generalization performance
  - Quantifying impact of ML on ultrasonic NDE probability of detection (POD)

- Methods for demonstrating confidence in ML performance being studied in other applications and as part of Standards development activities



ROC Curve Comparison for Defect/No Defect Classification

(From Cruz et al, Ultrasonics 73, pp 1-8 (2017))



Classification accuracy vs Data Imbalance

*Literature on ML application to other NDE techniques also shows promise though not all studies address the above factors

A Need for Representative, Common, Public NDE Data Sets

- Sample size and representativeness seem to be a limiting condition in most ML for ultrasonic NDE studies
  - Data augmentation approaches have been applied in some studies to mitigate sample size concerns
  - Unclear whether data augmentation helps with generalization performance
- Representative, common data sets
  - Enable comparison between methods
  - Support V&V approaches to demonstrate impact of ML on NDE reliability
  - Enable reproducibility of ML research results

Robustness of ML Solution

- Sensitivity studies relative to model parameters are likely to be important to improving confidence in the reported results
  - Impact of noise in the data on the results is part of the assessment
  - Model tuning should be a standard part of the methodology for developing ML solutions for NDE
- Effective V&V approaches to quantify confidence in ML solution necessary
- Robustness assessment/V&V of ML will need information on software tools and development environment
  - Enables assessment of potential limitations with tools
  - Increases reproducibility of results
  - Simplifies maintainability of code-base

Summary and Future Plans

- Assessment of literature demonstrates the potential of ML for automating ultrasonic NDE data analysis
  - Literature survey to assess the state of art in ML for ultrasonic NDE being finalized for publication
- Literature review identified several open questions related to the impact of factors that influence ML performance, and the contribution of ML to increasing NDE reliability
- Recommendations being formulated for addressing these questions and developing the technical bases to support regulatory decisions regarding reviews of relief requests and Code actions that that include ML
- Future plans: compilation of reference data sets and empirical studies to address open questions from literature review

Questions?