TIPS FOR EFFECTIVE MACHINE LEARNING IN NDT/E

BY JOEL B. HARLEY, SUHAIB ZAFAR, AND CHARLIE TRAN

The proliferation of machine learning (ML) advances will have long-lasting effects on the nondestructive testing/evaluation (NDT/E) community. As these advances impact the field and as new datasets are created to support these methods, it is important for researchers and practitioners to understand the associated challenges. This article provides basic definitions from the ML literature and tips for nondestructive researchers and practitioners to choose an ML architecture and to understand its relationships with the associated data. By the conclusion of this article, the reader will be able to identify the type of ML architecture needed for a given problem, be aware of how characteristics of the data affect the architecture's training, and understand how to evaluate the ML performance based on properties of the dataset.

Introduction

Advances in ML have consistently generated headlines in the past few years. These developments can be attributed to sophisticated algorithms, faster hardware, and reduced costs for data storage. The natural consequence of such advancements is the deluge of datasets, often known as the age of big data. ML algorithms, especially deep learning, capitalize on these foundations, finding applications in speech recognition and object detection while opening up new possibilities through innovations such as ChatGPT (OpenAI 2023). These applications vary considerably from one another, yet the main task in each case is to recognize patterns in datasets.

Pattern recognition is arguably the primary driving force behind new scientific and engineering discoveries. For instance, Kepler utilized the observations of Tycho Brahe in astronomy to derive the laws governing planetary motion, which formed the basis for classical mechanics (Bishop 2006). However, data was not a driving force behind scientific inquiry until recently (Brunton et al. 2020), and these trends have also impacted NDT/E (Taheri et al. 2022), with recent advances such as crack detection in concrete using neural networks (Saleem and Gutierrez 2021) or identifying damage modes in composite structures via clustering algorithms (Xu et al. 2020). Neural networks are one of the most widely used algorithms today and can be understood as a class of mathematical models inspired by the structure of the human brain.

However, utilizing neural networks, or ML in general, for tasks such as defect detection or aiding data interpretation is a familiar trend in NDT/E. Martín et al. (2007) published a study in 2007 to interpret ultrasonic oscillograms obtained via the pulse-echo method with the aid of neural networks. Even earlier, in the 1990s, Mann et al. (1992) presented the use of neural networks to classify ultrasonic signals obtained from microfiber cracking in a specimen built using a metal matrix composite. These examples demonstrate that the NDT/E community has long recognized the need to augment human judgment with pattern recognition algorithms.

Despite these advances, limitations of ML in NDT/E have mitigated its impact on the field when compared with other disciplines. A widely acknowledged problem is the limited amounts of data available, which is the driving force behind the success of ML in many applications. Even if the lack of training data is not an issue with data-intensive applications, such as acoustic emission testing (Sikorska and Mba 2008), acquiring data with a high signal-to-noise ratio (SNR) is a significant hurdle. Finally, an adequate level of understanding and experience in ML techniques is required to ensure the accurate performance of algorithms, which currently needs improvement (Vejdannik et al. 2019).

In this article, we address important challenges in applying ML to NDT/E by providing guidelines for practitioners and researchers on building high-quality datasets and using appropriate algorithms to ensure high performance from trained ML models. The desired outcome of this effort is to encourage progress in realizing the full potential of ML in NDT/E, leading to more accurate and efficient testing methods in the future. Note that the focus of this article is on how to assess datasets and results. Detailed descriptions of the ML algorithms can be found in other papers (Taheri and Zafar 2023).

Forms of Machine Learning

ML can be divided into various learning paradigms, each with its characteristics and uses. Below are descriptions for two of these paradigms: supervised learning and unsupervised learning. Examples of supervised learning and unsupervised learning are illustrated in Figure 1.

Supervised learning: An ML paradigm that trains the parameters (often numerical weights) of a model from input data (features) and known output data (labels). Supervised learning is the most popular ML paradigm due to the ease at which model training can be directly translated to the target task. The key element of supervised learning is the availability of labeled data. Yet in

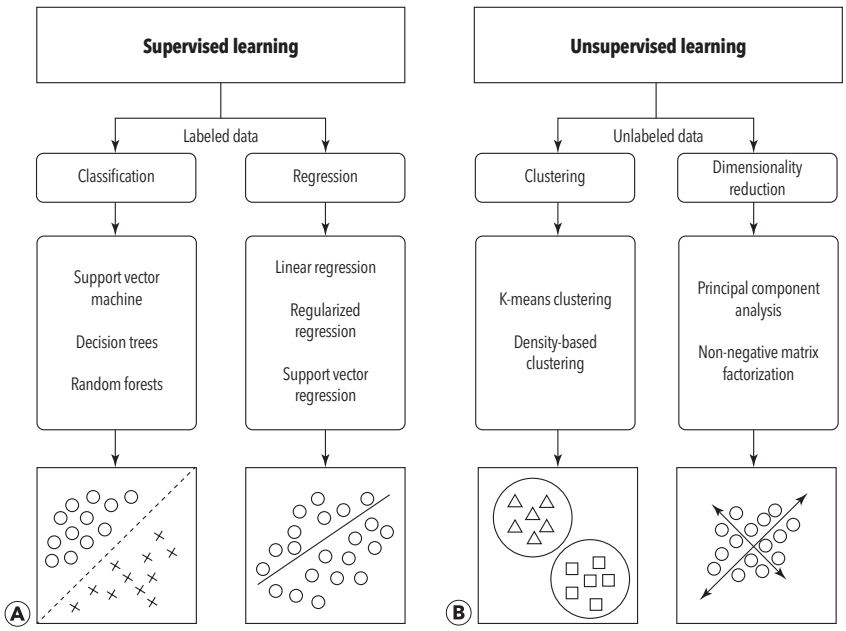


Figure 1. Learning paradigms of machine learning: (a) supervised learning algorithms utilize labeled data, which allows algorithms to be trained directly on the downstream task (classification and regression); (b) unsupervised algorithms utilize unlabeled data, which are primarily used for clustering and dimensionality reduction. Semi-supervised learning algorithms incorporate characteristics of both of these paradigms.

NDT/E, obtaining reliable labels is a specialized and time-consuming task, which is further complicated as manual delineation of discontinuities introduces user subjectivity. In turn, mislabeled data can be counteractive to the learning process of supervised learning ML models (Taheri and Zafar 2023; Lever et al. 2017).

Unsupervised learning: An ML paradigm from which an ML model is trained from the input data features, but without known data labels. Clustering is one of the most well-known forms of unsupervised learning, wherein data is divided into discrete groups. Furthermore, dimensionality reduction and manifold learning methods such as principal component analysis (PCA) (Lever et al. 2017; Yang et al. 2022) and t-distributed stochastic neighbor embedding (tSNE) (van der Maaten and Hinton 2008) are forms of unsupervised learning. Unsupervised learning is useful in NDT/E due to the challenges of obtaining labels.

Tips: If dependable labels can be obtained for a dataset, a supervised learning paradigm is often the simplest and most accurate. Assuming no labels are known, unsupervised learning is powerful but requires domain-specific insights from the user. Unsupervised learning also generally lacks metrics for standardized evaluation.

Types of Learning Tasks

Each ML paradigm can take on different tasks. In this subsection, we subdivide supervised learning into its two most common tasks (classification and regression) and subdivide unsupervised learning into its two most common tasks (clustering and dimensionality reduction). These subgroups are illustrated in Figure 1.

Classification: A supervised ML model performs classification when it determines if the input data belongs to one of a discrete set of "classes," or categories. For example, different defect types (e.g., delamination, crack, no defect) may represent different classes that we may observe.

Regression: A supervised ML machine model performs regression when estimating the value of a continuous dependent variable from an input independent variable. For example, an ML model may process imaging NDT/E data to estimate the size of a defect.

Clustering: The clustering task aims to classify data without known information by identifying groups, or clusters, of data that are similar to each other in some manner. Clustering can be valuable for identifying unknown relationships between the data, such as the presence of outlier data that could correspond to a discontinuity.

Dimensionality reduction: The aim of dimensionality reduction is to reduce the data into its essential features. Many compression and denoising algorithms can be considered forms of dimensionality reduction (Yang et al. 2022). It can separate components (e.g., multiple reflections from an ultrasonic B-scan) that reconstruct the data when added together (Liu et al. 2015). This is sometimes referred to as blind source separation.

Tips: It is important to determine the appropriate learning task for a given problem as it dictates the choice of an ML model and the associated challenges. Figure 1 describes the most common ML models used for each task.

Characteristics of Machine Learning Datasets and Architectures

Most ML architectures learn only from the provided data. As a result, ML model performance is highly dependent on the dataset quality. The classic bias-variance tradeoff is one of the most common challenges we must consider when building a dataset and choosing an architecture.

Bias: One of the most significant issues that one must consider when creating a dataset is to consider the inherent bias that the dataset exhibits and how it affects the ML model. That is, a dataset will be biased if the training data (i.e., the input data and labels that are used to initially train the model) tends to better represent one scenario over another (Mehrabi et al. 2022). Note that bias is not inherently bad since you may want to focus on a particular scenario (Miceli et al. 2022), but it is important to acknowledge that bias. For example, an ML model trained

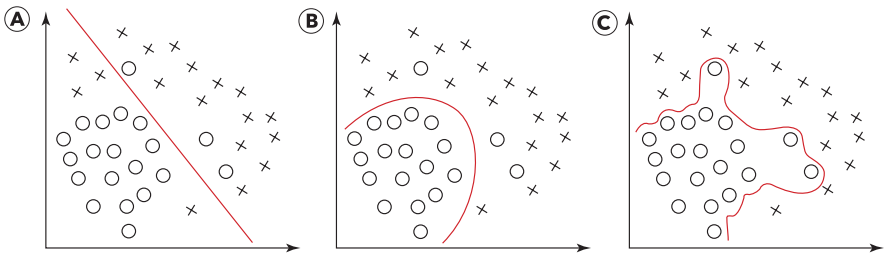


Figure 2. Model fitting: (a) underfitting; (b) ideal fitting; and (c) overfitting. An underfitting model characteristically suffers from poor performance in the training data, being unable to learn the relationships within the data. On the other hand, an overfitting model characteristically suffers from over-performing on the training data (often viewed as "memorization") and fails to generalize onto new data samples. Thus, a fundamental goal of machine learning algorithms is to find an ideal fitting.

on simulated data (for which we can produce an abundance of labeled data) will learn the specific characteristics of the simulated data, but it may not also represent experimentally measured data. If the labels are imbalanced (e.g., there are twice as many cracks as delaminations), then the data will be inherently more likely to predict the larger class. In short, if a characteristic of our data is imbalanced (e.g., twice as many measurements originate from aircraft wings than bridges), then the predictions will be more accurate for those dominant characteristics. An underfit ML model is created when trained with a biased dataset or when the ML model has too few parameters (Figure 2). Such a model fails to learn specific characteristics from the data, leading to poor performance (the classic bias-variance tradeoff is illustrated in Figure 3).

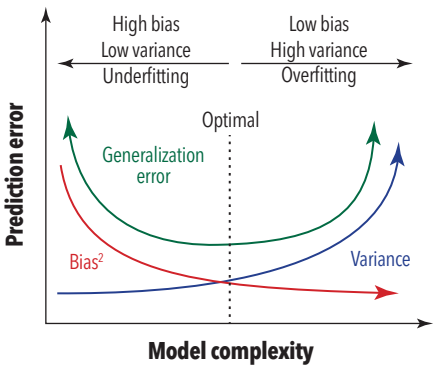


Figure 3. Bias-variance tradeoff curve. Machine learning models strive to balance bias and variance. Simple machine learning models typically have fewer parameters, wherein the high bias and low variance are characteristic of model underfitting. On the other hand, complex machine learning models have a large number of parameters, wherein the low bias and high variance are characteristic of model overfitting.

Variance: The effects of data imbalances are difficult to gauge in part due to the variance in the dataset, another factor that must be considered when building data. A common question posed by non-ML practitioners is often "How much data do you need?" The answer is usually "it depends" due to the inherent variance in the input data. For example, if a crack looks identical in every single measurement, then the dataset has very low variance. In this scenario, you may not need a learning system because one datum of a crack sufficiently describes all other examples (although some pattern recognition is still necessary). In contrast, if there are a million different and unique permutations of how a crack is represented, then the ML model will need at least a million examples to correctly classify cracks. In reality, there are usually complex relationships between all data corresponding to cracks, which the ML model can learn. A highly variable dataset with too few training examples and too many parameters to learn can yield an overfit ML model (Figure 2). Such a model may find uninformative relationships in noise, leading to poor performance (Figure 3) (Belkin et al. 2019).

Interpretability: One should also consider the interpretability of an ML architecture. An interpretable ML model is one from which humans can comprehend how a decision is made (Du et al. 2019). In general, there is a negative correlation between accuracy and model interpretability (Figure 4). Gaining interpretability is a difficult problem due to the nature of black-box models, non-linearities, and high-dimensional data visualizations. Deep neural networks are the prime example, being the most accurate models but with little to no interpretability of the model decision-making. On the other hand, linear models (e.g., linear regression) are very interpretable, yet often less accurate.

Tips: Misunderstanding bias and variance is a significant pitfall for early ML practitioners. For example, novice deep learning practitioners often default toward increasing the number of layers in a neural network, thereby increasing the model complexity. However, such an architecture is not only more computationally demanding but can in some cases be less effective (due to overfitting) and less interpretable than a simpler architecture. For this reason, deep neural networks are unfavorable in situations with limited data samples of potentially high variance and situations where interpretability and accountability are important. In such a scenario, users may often analyze their problem using conventional ML models, such as support vector machines or linear regression

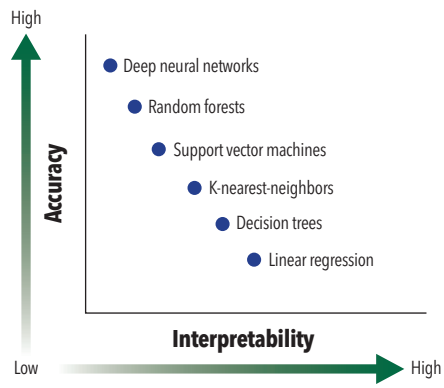


Figure 4. Model accuracy versus interpretability. In machine learning, increased accuracy has a natural consequence of decreased interpretability. Accurate models tend to capture nonlinear and non-smooth relationships, while interpretable models tend to capture linear and smooth relationships.

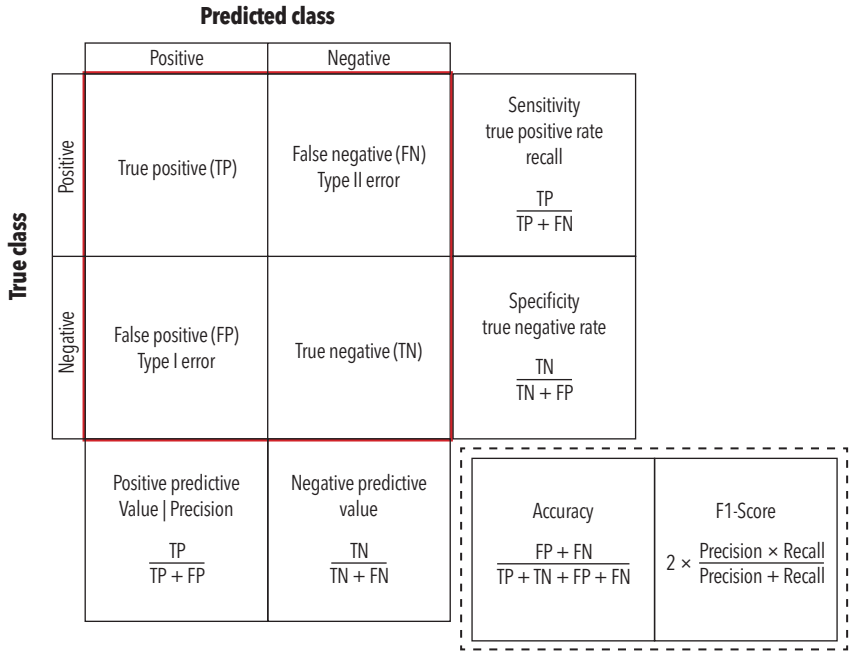


Figure 5. A confusion matrix is used to evaluate the performance of a classifier, summarizing the information between true and predicted classifications. The confusion matrix entails the number of true positives, false negatives, false positives, and true negatives. Further classification metrics may be extracted (e.g., sensitivity, specificity, accuracy, etc.) to measure different aspects of the classifier.

models, which are generally more interpretable (Figure 4). In essence, applying a model architecture should be inspired by the data and underlying factors at hand, especially for new datasets that have not been utilized for ML in the past.

Metrics for Evaluation

Evaluation metrics are performance measures for comparing ML models and understanding specific characteristics of the data or task. This is in part due to the bias and variance within the data. In particular, different evaluation metrics can be used to attain either a holistic performance or a class-specific measure. Here we review several of the most widely used metrics for evaluating the performance of ML models.

Confusion matrix: The confusion matrix visualizes the predicted values against the true values. Elements on the diagonal of the matrix indicate the number of true predictions of the model to the true class (true positives and true negatives). The off-diagonal elements indicate incorrect predictions. Reading the confusion matrix tends to give further insight as to what types of errors are made for a model and allows a holistic set of evaluation metrics. We provide a typical illustration in Figure 5, together with the common name of such evaluation metrics. The confusion matrix need not be binary but can be conducted in a multi-class fashion. However, in the multi-class scenario, summarizing the model performance may be cumbersome, and traditionally each class is evaluated in a one-versus-all manner.

Accuracy: Accuracy is often the most common evaluation metric. The accuracy is the proportion of the model predictions correct relative to the true class. From the perspective of the confusion matrix, this is equivalent to the sum of the diagonal divided by the sum of all of the values. Accuracy is an easy value to understand. However, for imbalanced datasets, the accuracy can be uninformative. For example, a common scenario in NDT/E might be that 99% of the data is from a normal material and 1% of the data is a material with a discontinuity. If 100% of the data is classified as normal, then the accuracy is 99%. This is often considered a good result until you recognize that none of the discontinuities are identified.

Recall: Also known as sensitivity or the true positive rate (TPR), the recall is the proportion of true positive cases that are correctly predicted. In binary classification, notice that if 99% of the labels do not correspond to the class of interest, and 100% of the predictions correspond to those classes, then the recall will be 0. Hence, recall can be suitable when data is imbalanced.

Precision: Also known as the positive predictive value (PPV), measures the proportion of correct positive predictions made. Observe, if 99% of the labels do not correspond to the class of interest, and 100% of the predictions correspond to those classes, then the precision will be 0. Therefore, precision can be advantageous when data is imbalanced.

F1 score: The F1 score is a metric designed to summarize both precision and recall. It is defined as the harmonic mean of precision and recall. The harmonic mean, as opposed to the arithmetic mean, addresses large deviations between precision and recall. For example, if the precision for a class is 0, and the recall is 1, then the arithmetic mean evaluates to 0.5, which may naively indicate a random classifier. On the other hand, the harmonic mean in this scenario equates to 0, revealing the classifier is predicting only one class.

Receiver operating characteristic curve: The receiver operating characteristic (ROC) curve can be generated when the confusion matrix varies as a function of a set call criterion (Figure 6). This metric originates from traditional statistical hypothesis testing in which a binary classifier is based upon the premise that some statistic is above or below a threshold. In a binary classification scenario, the ROC curve shows the false positive rate versus the true positive rate for all threshold values. To summarize the ROC, the area under the ROC curve (AUC) is often reported, where a perfect classifier attains a value of 1 and a random classifier attains an AUC of 0.5. The AUC metric is valuable as it is invariant of the chosen threshold

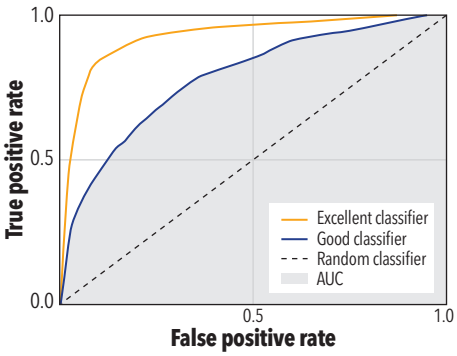


Figure 6. Receiver operating characteristic (ROC) curve. The ROC curve is achieved by plotting the false positive rate versus the true positive rate at each classification threshold. The quality of the ROC curve can be summarized by the area under the curve (AUC) shaded in gray.

and therefore evaluates the overall classifier rather than some user-chosen value. The AUC is also a feasible metric for imbalanced data.

Tips: Careful choice of evaluation metrics should be selected based upon the bias and variance of the dataset. For an unbiased, well-balanced dataset, accuracy is often the most characteristic of the model performance. In NDT/E, we are often concerned with the true positive rate, which is also known as the probability of (defect) detection or the recall of a defect. In other NDT/E scenarios, we may want to ensure that normal materials are not predicted as material defects (e.g., delaminations), in which case, the false call rate (also known as the false negative rate) or the precision score may be more valuable. Note the true positive and false positive rates are utilized in traditional NDT/E probability of detection assessment (Cherry and Knott 2022). In the cases where we want a balance between the recall and precision scores, the F1 score becomes a valuable metric.

Conclusion

ML has a significant potential to contribute to the NDT/E community. However, successful usage of ML algorithms demands greater insight into their capabilities and intricacies. This sentiment is also true for those in the community building new datasets for ML practices. Understanding the basic capabilities of ML paradigms, navigating how bias and variance within the data affect the ML model, and establishing how performance will be measured will help the community create datasets that have the greatest impact. ME

ACKNOWLEDGMENTS

Work on this paper is partially funded by the United States Air Force contract FA8650-18-C-5015.

AUTHORS

Joel B. Harley: Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611

Suhaib Zafar: Stellantis Chrysler Technology Center, Auburn Hills, MI 48326

Charlie Tran: Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611

CITATION

Materials Evaluation 81 (7): 43–47
https://doi.org/10.32548/2023.me-04358

REFERENCES

Belkin, M., D. Hsu, S. Ma, and S. Mandal. 2019. "Reconciling modern machine-learning practice and the classical bias-variance trade-off." *Proceedings of the National Academy of Sciences of the United States of America* 116 (32): 15849–54. https://doi.org/10.1073/pnas.1903070116.

Bishop, C. M. 2006. *Pattern Recognition and Machine Learning*. Springer New York.

Brunton, S. L., B. R. Noack, and P. Koumoutsakos. 2020. "Machine Learning for Fluid Mechanics." *Annual Review of Fluid Mechanics* 52 (1): 477–508. https://doi.org/10.1146/annurev-fluid-010719-060214.

Cherry, M., and C. Knott. 2022. "What is probability of detection?" *Materials Evaluation* 80 (12): 24–28. https://doi.org/10.32548/2022.me-04324.

Du, M., N. Liu, and X. Hu. 2019. "Techniques for interpretable machine learning." *Communications of the ACM* 63 (1): 68–77. https://doi.org/10.1145/3359786.

Lever, J., M. Krzywinski, and N. Altman. 2017. "Principal component analysis." *Nature Methods* 14 (7): 641–42. https://doi.org/10.1038/nmeth.4346.

Liu, C., J. B. Harley, M. Bergés, D. W. Greve, and I. J. Oppenheim. 2015. "Robust ultrasonic damage detection under complex environmental conditions using singular value decomposition." *Ultrasonics* 58:75–86. https://doi.org/10.1016/j.ultras.2014.12.005.

Mann, L. L., T. E. Matikas, P. Karpur, and S. Krishnamurthy. 1992. "Supervised backpropagation neural networks for the classification of ultrasonic signals from fiber microcracking in metal matrix composites." in *IEEE 1992 Ultrasonics Symposium Proceedings*. Tucson, AZ. https://doi.org/10.1109/ULTSYM.1992.275983.

Martín, Ó., M. López, and F. Martín. 2007. "Artificial neural networks for quality control by ultrasonic testing in resistance spot welding." *Journal of Materials Processing Technology* 183 (2–3): 226–33. https://doi.org/10.1016/j.jmatprotec.2006.10.011.

Mehrabi, N., F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. 2022. "A Survey on Bias and Fairness in Machine Learning." *ACM Computing Surveys* 54 (6): 1–35. https://doi.org/10.1145/3457607.

Miceli, M., J. Posada, and T. Yang. 2022. "Studying Up Machine Learning Data: Why Talk About Bias When We Mean Power?" *Proc. ACM Hum.-Comput. Interact.* 6: 1–14. https://doi.org/10.1145/3492853.

OpenAI. 2023. "GPT-4 Technical Report." arXiv:2303.08774. https://doi.org/10.48550/arXiv.2303.08774.

Saleem, M., and H. Gutierrez. 2021. "Using artificial neural network and non-destructive test for crack detection in concrete surrounding the embedded steel reinforcement." *Structural Concrete* 22 (5): 2849–67. https://doi.org/10.1002/suco.202000767.

Sikorska, J. Z., and D. Mba. 2008. "Challenges and obstacles in the application of acoustic emission to process machinery." *Proceedings of the Institution of Mechanical Engineers. Part E, Journal of Process Mechanical Engineering* 222 (1): 1–19. https://doi.org/10.1243/09544089JPME111.

Taheri, H., and S. Zafar. 2023. "Machine learning techniques for acoustic data processing in additive manufacturing in situ process monitoring — A review." *Materials Evaluation* 81 (7): 50–60.

Taheri, H., M. Gonzalez Bocanegra, and M. Taheri. 2022. "Artificial Intelligence, Machine Learning and Smart Technologies for Nondestructive Evaluation." *Sensors (Basel)* 22 (11): 4055. https://doi.org/10.3390/s22114055.

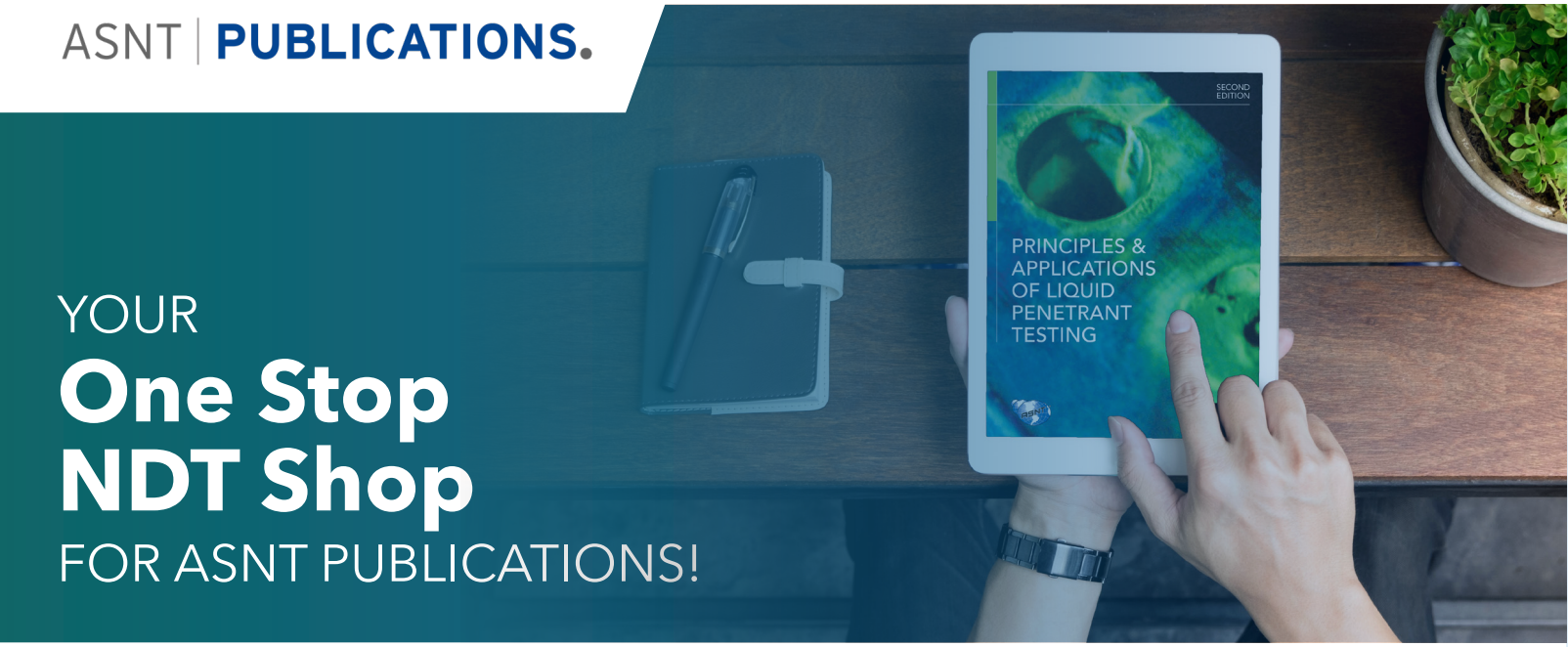
van der Maaten, L., and G. Hinton. 2008. "Visualizing Data using t-SNE." *Journal of Machine Learning Research* 9 (86): 2579–605.

Vejdannik, M., A. Sadr, V. H. C. de Albuquerque, and J. M. R. S. Tavares. 2019. "Signal Processing for NDE," in *Handbook of Advanced Nondestructive Evaluation*. eds. N. Ida and N. Meyendorf. Springer. pp. 1525–1543. https://doi.org/10.1007/978-3-319-26553-7_53.

Xu, D., P. F. Liu, Z. P. Chen, J. X. Leng, and L. Jiao. 2020. "Achieving robust damage mode identification of adhesive composite joints for wind turbine blade using acoustic emission and machine learning." *Composite Structures* 236:111840. https://doi.org/10.1016/j.compstruct.2019.111840.

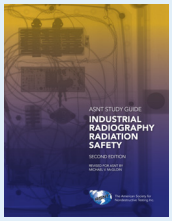
Yang, K., S. Kim, and J. B. Harley. 2022. "Guidelines for effective unsupervised guided wave compression and denoising in long-term guided wave structural health monitoring." *Structural Health Monitoring*. https://doi.org/10.1177/14759217221124689.

ASNT | **PUBLICATIONS.**



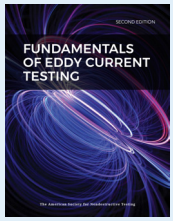
YOUR
One Stop NDT Shop
FOR ASNT PUBLICATIONS!

NEW PUBLICATIONS



Industrial Radiography Radiation Safety Study Guide
second edition

Order 10107 | ebook 10107-e

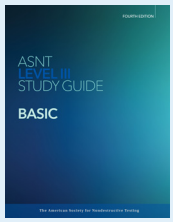


Fundamentals of Eddy Current Testing
second edition

Order 10133 | ebook 10133-e

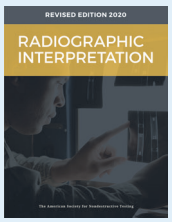
A Guide to Personnel Qualification & Certification
fifth edition

Order 10178 | ebook 10178-e



ASNT Level III Study Guide: Basic
fourth edition

Order 10108 | ebook 10108-e



Radiographic Interpretation
revised edition 2020

Order 10168 | ebook 10168-e

Quick Reference Method Cards
These cards (sold individually or as a set) help both the novice and experienced inspector by providing examples of testing techniques and sample calculations.

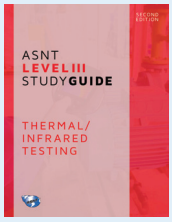
Order 10144 | ebook 10144-e

ASNT Level II Study Guides
Available in PT, RT, VT, MT, and UT

Liquid Penetrant Testing
third edition

Order 6101
ebook 6101-e

ASNT Level III Study Guides
Available in Basic, PT, RT, VT, MT, UT, IR, LT, and ET



Thermal/Infrared Testing
second edition

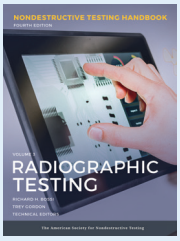
Order 2265
ebook 2265-e

Standards
Standards, recommended practices, and supplementary materials

Recommended Practice No. SNT-TC-1A: Personnel Qualification and Certification in Nondestructive Testing (2020)

Order 2073 | ebook 2073-e

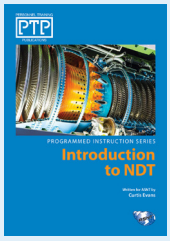
Nondestructive Testing Handbooks
The NDT Handbook covers all major methods of nondestructive testing, each major method in its own volume



Nondestructive Testing Handbook, Radiographic Testing
fourth edition, volume 3

Order 0144 | ebook 0144-e

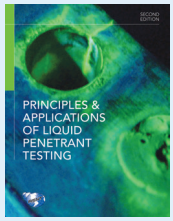
PTP Programmed Instruction Series
The PTP PI Series is a self-study resource for Level I and II candidates



Introduction to NDT

Order 1530
ebook 1530-e

NDT Method References
Resources for learning and reviewing the fundamentals of NDT methods



Principles and Applications of Liquid Penetrant Testing
second edition

Order 2204 | ebook 2204-e

NEW! Individual Handbook Chapters
Browse ebooks.asnt.org.

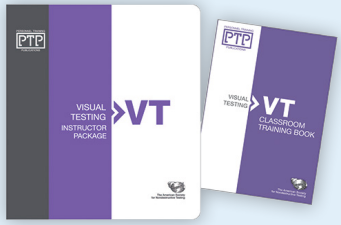
NDT Handbook Vol. 1: Liquid Penetrant Testing
15 chapters available

NDT Handbook Vol. 2: Leak Testing
15 chapters available

NDT Handbook Vol. 3: Radiographic Testing
17 chapters available

Aerospace NDT Industry Handbook
19 chapters available

Personnel Training Publications (PTP) Classroom Training Series
The PTP Series is available in six methods: PT, MT, UT, ET, RT, and VT
Each volume covers Level I and II



Visual Testing (VT) Instructor Package and Student Package
second edition

Instructor Package
Order 1655

Student Package
Order 1665

Q&As
Level I, II, and III sample questions for instruction and exam preparation

Ultrasonic Testing (UT) Questions & Answers Set: Method and Techniques
Order 1001K | ebook 1001K-e

From compelling research to the knowledge you need to earn or maintain your certifications, discover what's new in ASNT's catalog of publications. Select publications are available in both print and digital versions.

SHOP NOW at **asnt.org/store**.

ASNT...CREATING A SAFER WORLD!®

MACHINE LEARNING TECHNIQUES FOR ACOUSTIC DATA PROCESSING IN ADDITIVE MANUFACTURING IN SITU PROCESS MONITORING A REVIEW

HOSSEIN TAHERI* AND SUHAIB ZAFAR†

ABSTRACT

There have been numerous efforts in the metrology, manufacturing, and nondestructive evaluation communities to investigate various methods for effective in situ monitoring of additive manufacturing processes. Researchers have investigated the use of a variety of techniques and sensors and found that each has its own unique capabilities as well as limitations. Among all measurement techniques, acoustic-based in situ measurements of additive manufacturing processes provide remarkable data and advantages for process and part quality assessment. Acoustic signals contain crucial information about the manufacturing processes and fabricated components with a sufficient sampling rate. Like any other measurement technique, acoustic-based methods have specific challenges regarding applications and data interpretation. The enormous size and complexity of the data structure are significant challenges when dealing with acoustic data for in situ process monitoring. To address this issue, researchers have explored and investigated various data and signal processing techniques empowered by artificial intelligence and machine learning methods to extract practical information from acoustic signals. This paper aims to survey recent and innovative machine learning techniques and approaches for acoustic data processing in additive manufacturing in situ monitoring.

KEYWORDS: additive manufacturing, in situ monitoring, acoustic, machine learning, data processing

Introduction

Various additive manufacturing (AM) methods are utilized for manufacturing parts with complex geometries and complicated features that are either unfeasible or highly challenging to produce via traditional manufacturing techniques. This outstanding capability of AM provides substantial design flexibility and facilitates the production of complex parts with marginal added cost compared to subtractive and traditional manufacturing methods (Calta et al. 2018). Laser powder bed fusion (LPBF), directed energy deposition (DED), and wire arc additive manufacturing (WAAM) are among the most popular methods of metal AM (Koester et al. 2018). Fused deposition modeling (FDM), stereolithography (SLA), direct ink writing (DIW), and selective laser sintering (SLS) are the most common AM techniques for polymers (Baechle-Clayton et al. 2022; Lee et al. 2020).

The AM processes not only can cause different mechanical properties for the parts manufactured, but also lead to the potential generation of specific types of discontinuities and defects in AM parts (Koester et al. 2018, 2019b; Taheri et al. 2017). The types of defects in AM parts significantly depend on manufacturing process conditions and type of materials. A summary of defect types, causes of defect generation, and their potential effect on AM parts is presented in Table 1.

Although inspection and quality assessment for the manufactured parts can be done after the production is finished (ex situ), there are several significant challenges in traditional ex situ inspection methods. One of the major challenges of traditional inspection of AM parts is due to the capability of AM techniques to produce complex-geometry components. This is an outstanding capability for AM but makes traditional inspection of AM parts extremely challenging since many available nondestructive testing (NDT) techniques have been developed for simpler geometries (Bond et al. 2019). Another primary concern in post-production or ex situ inspection of AM parts is that AM techniques are used to manufacture many critical, high-valued, or exotic parts. Possible rejection of such unique parts due to unacceptable quality causes a significant loss of time and cost and is not a desirable outcome for industries (Koester et al. 2018c; Taheri 2018). Despite the complexity of the processes in AM, the layer-by-layer deposition of materials allows the measurement and recording of large amounts of data on each layer for statistical process monitoring and quality assessment (Grasso and Colosimo 2017; Koester et al. 2018b).

* The Laboratory for Advanced Non-Destructive Testing, In-situ Monitoring and Evaluation (LANDTIE), Department of Manufacturing Engineering, Georgia Southern University, Statesboro, GA, USA 30458; htaheri@georgiasouthern.edu
† Stellantis, Chrysler Technology Center, 800 Chrysler Dr., Auburn Hills, MI 48326, USA

Materials Evaluation 81 (7): 50–60
https://doi.org/10.32548/2023.me-04356