RESEARCH ARTICLE | MAY 08 2019

Intelligence augmentation in nondestructive evaluation FREE

John C. Aldrin ; Eric A. Lindgren; David S. Forsyth

AIP Conf. Proc. 2102, 020028 (2019)
https://doi.org/10.1063/1.5099732

Intelligence Augmentation in Nondestructive Evaluation

John C. Aldrin[2, a)], Eric A. Lindgren[1], David S. Forsyth[3]

[1]*Air Force Research Laboratory (AFRL/RXCA), Wright-Patterson AFB, OH 45433, USA*
[2]*Computational Tools, 4275 Chatham Ave., Gurnee, IL 60031, USA*
[3] *TRI/Austin, Austin, TX 78746, USA*

[a)] Corresponding author: aldrin@computationaltools.com

Abstract. In recent years, advances have been made in the field of machine learning and artificial intelligence (AI), primarily through developments in deep learning neural networks (DLNN). Challenges however exist with transitioning emerging DLNN algorithms directly for NDE applications. As a counterpoint to AI, intelligence augmentation (IA) refers to the effective use of information technology to enhance human intelligence. While attempting to replicate the human mind has encountered many obstacles over the years, IA has a much longer history of success. All forms of information technology, from writing cuneiform on clay tables to computers and smartphones today, have essentially been developed to enhance the information processing capabilities of the human mind. This paper introduces a series of best practices for intelligence augmentation in NDE, highlighting how the operator should interface with NDE data and algorithms. Algorithms clearly have a great potential to help alleviate the burden of 'big data' in NDE; however, it is important that operators are involved in both secondary indication review, and the detection of rare event indications not addressed well by typical algorithms. Several past examples of transitioning algorithms for NDE applications are presented, emphasizing the successful interfacing of operator and software for optimal data review and decision making.

INTRODUCTION

There is a continuing need to develop nondestructive methods for the detection of damage and the characterization of the material state. With improved capability, there is great potential to enable and refine methods to certify new materials, sustain existing systems, and enhance methods to forecast future maintenance of engineered systems [1,2]. The subject of NDE characterization is a very broad topic, with a long history in the context of quantitative nondestructive evaluation [3]. In recent years, advances in NDE sensor technology, scanning hardware, imaging capabilities, and data analysis algorithms have provided improvements in NDE sensitivity, resolution and coverage. Advanced manufacturing for composite materials and additive manufacturing (AM) technology have provided new challenges but also opportunities for enhancing NDE capability at multiple stages of manufacturing process.

For the USAF, there is a long and on-going effort to both develop and implement automation for sustainment, to ensure NDE reliability and help mitigate maintenance time and cost [1,4]. Some examples of NDE automated system for aerospace applications are shown in Fig. 1, covering a range of techniques (eddy current, ultrasound) and a variety of environments, including manufacturing, the depot, and field-service, sometimes with limited accessibility. Going forward, there are additional opportunities to better leverage NDE data, through such initiatives as the Digital Thread and Digital Twin [5,6]. The Digital Thread provides a means to track all digital information regarding the manufacturing and sustainment of a component and system. Through digitizing the entire inspection and the quantitative evaluation of structural performance, this capability will drastically speed the material review board (MRB) process for evaluating instances of non-conformance [7]. As well, data capture enables quantitative evaluation of the variance between the design and the as-manufactured (and as-maintained) parts [6]. Similarly, the Digital Twin concept provides a digital equivalent of a system in its present state, and then exercises the digital twin model through various use scenarios to determine how it will perform and forecast possible emerging maintenance issues. These initiatives require comprehensive programs for digital data capture and registration with part models, that are frequently missing for many NDE inspections today [8]. As well, special algorithms and processes are needed to subsequently convert NDE data into information of practical use for maintainers [9].

Analogous in many respects to these USAF initiatives, Industry 4.0 is a term developed by German industry leaders and researchers to describe how the Internet of Things (IoT) will improve engineering, manufacturing, logistic and life cycle management processes [10]. The number 4.0 refers to a 4th industrial revolution. Beginning in the 1700s, three major waves of technological changes transformed the industrial landscape and increased productivity: (1) mechanization and water/steam power, (2) mass production (e.g. assembly lines) and electricity, and (3) computers

45th Annual Review of Progress in Quantitative Nondestructive Evaluation, Volume 38
AIP Conf. Proc. 2102, 020028-1–020028-10; https://doi.org/10.1063/1.5099732
Published by AIP Publishing. 978-0-7354-1832-5/$30.00

and automation. The 4th industrial revolution is expected to be based on connected cyber-physical systems. There is a parallel vision for the next-generation of NDE capability referred to as NDT 4.0 [11,12]. The most important aspect of the integration of NDT 4.0 is leveraging automation in the evaluation of the work piece under consideration and connecting the part state with life cycle management.

At the heart of NDE automation and cyber-physical systems are algorithms. A survey of algorithms in NDE, including case studies of USAF applications, was recently presented in ref. [13]. In recent years, promising advances have been made in the field of machine learning (ML) and artificial intelligence (AI), primarily through developments in deep learning neural networks (DLNN) [14-17]. However, challenges exist with transitioning emerging DLNN algorithms directly for NDE applications. Training deep learning neural networks requires very large, well-understood data sets, which is frequently not available for NDE applications. As well, there are significant concerns about the reliability and adaptability of such algorithms to completely perform complex NDE data review tasks. In addition, a critical component of any tool is the interface with the human that uses it. Care must be taken with the implementation of automation to ensure that the operator has the necessary awareness and control as needed. The focus of this paper is this emerging interface between NDE hardware, software and algorithms and the human inspectors and engineers. Experience and perspective on the transition of algorithms for NDE application will be addressed, highlighting emerging trends and challenges with achieving the vision of Industry 4.0.

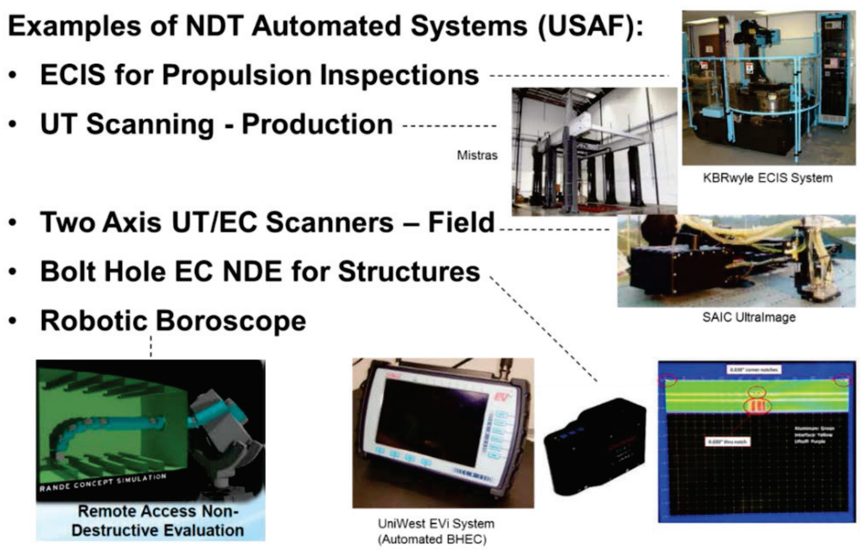

FIGURE 1. Examples of NDT automated systems for aerospace applications.

EMERGING ALGORITHMS AND NDE: PROS AND CONS

Three classes of algorithms for NDE detection and characterization can be organized as follows: (1) algorithms based on NDE expert knowledge and procedures (heuristic algorithms), (2) model-based inversion, (3) algorithms incorporating statistical classifiers and/or machine learning. The most basic algorithm is one based on human experience. The term *heuristic algorithm* is useful to describe a class of algorithm, based on learning through discovery, incorporating 'rules-of-thumb', 'common sense', and 'practical knowledge'. This first class of algorithms essentially encodes all key evaluation steps and criteria used by operators as part of a procedure within the algorithm. The second class of algorithm is *model-based inversion* that uses a 'first-principles' physics-based model with an iterative scheme to solve characterization problems. This approach requires accurate forward models and iteratively compares the simulated and measurement data, adjusting the model parameters until agreement (using an error metric) is reached. The third class of algorithm covers *statistical classifiers and machine learning* that are built through the fitting of a model function using measurement 'training' data with known states. Statistical representation of data classes can be accomplished using either frequentist procedures or Bayesian classification. Machine learning and artificial intelligence are general terms for the process by which computer programs can learn. Early work on machine learning built on emulating neurons through functions, as artificial neural networks using layered algorithms and a training process that mimics a network of neurons [14]. In recent years, impressive advances have been made in the field of machine learning, primarily through significant developments in deep learning neural network (DLNN) algorithms [15-17]. With the rapid increase in both computing power and large structured data sets, Geoffrey Hinton, Yann LeCun, and others began to improve neural network performance using many hidden layers, achieving several breakthroughs on test problems [16]. Large sets of high quality, well-characterized data have been critical for successful training of DLNNs. As well, software tools have been developed for training neural networks that better leverage advances in high performance computing (HPC). A recent overview on algorithms for NDE classification is summarized in reference [13]. These three categories could be extended to hybrid algorithms that contain mixtures of heuristic models, physical models and/or statistical/AI models. For example, John Launchbury, a director at DARPA, has described a third wave of artificial intelligence, beyond expert systems and statistical learning, "contextual adaption, which involves constructing reliable, explanatory models for real-world phenomena using sparse data, like humans do" [18,19]. This vision may be possible through building hybrid algorithms incorporating statistical models with expert and/or physical models that can more directly provide context.

There are a number of advantages with incorporating algorithms as part of an NDE technique. (A summary of the '*pros'* is presented in Table 1.) First, algorithms are typically very good at performing laborious and repetitive tasks. For most parts under test, either in manufacturing or in-service, the presence of critical NDE indications is a fairly rare event. So, the data review process can often be tedious task for most operators, expecting mostly good parts. Second, given the amount and complexity of some data review tasks performed for some inspections, such tasks can be a challenge, especially for inexperienced inspectors or inspections that are rarely performed. This trend appears to be growing with the increasing quantity of data acquired with automated scanning and array sensing systems. Third, in many instances, algorithms can perform the data review task faster than manual review, providing potential saving in maintenance time and cost. Fourth, algorithms are typically not biased by expectation, such as the frequency of indications in past inspections. With a reduction in errors, the overall risk of maintaining a component can be achieved. Fifth, algorithms can be designed in such a way to support the operator as a 'digital assistant'. Algorithms could potentially help alleviate the burden of 'mostly good data' and have operators focus on key data review tasks. As well, algorithms can be used to reduce the size and dimensionality of NDE data and present the operator with a reduced feature set for manual classification. Lastly, there are challenges with the aging workforce and transitioning expert knowledge to the next generation. Algorithms, if designed properly, can potentially be repositories for expert knowledge of an NDE organization.

While there is great promise with the application of NDE algorithms, there are a number of potential disadvantages with algorithm-based solutions to NDE inspection problems. (A summary of the '*cons*' is presented in Table 1.) First, the development and validation of reliable algorithms for NDE can be expensive. Training deep learning neural networks requires very large, well-understood data sets, which is frequently not readily available for NDE applications. Relative to many problem spaces like image, voice, and text recognition, NDE is considered 'data starved'. While the NDE community often possesses a large amount of data, the material state behind the data is often not perfectly known. Acquiring data from parts with well characterized damage states, such as cracks, corrosion, impact damage, etc., requires either high resolution NDE techniques (for example, microfocus computed tomography) for finger-printing, or destructive characterization for full verification. Thus, if training and validation data is not readily available, the manufacture of a large set of expensive test specimens with well-controlled flaw conditions will be required. The design, training and validation of algorithms also requires unique skills and many man-hours of engineering labor to successfully implement.

Second, algorithms typically perform poorly for scenarios that they are not trained to interpret. There have been concerns for decades about the reliability and adaptability of machine learning algorithms to completely perform complex NDE data review tasks. In NDE, early promising demonstrations have been performed by the NDE research community, but frequent issues concerning overtraining and robustness to variability for practical NDE measurement 'outside of the laboratory' have been noted [13]. Prior successful NDE applications of neural networks have been dependent on taking care to reduce the dimensionality of the data and provide reliable features as inputs for classification. Unfortunately, training deep learning neural networks requires very large, well-understood data sets relative to conventional classifier training. Designing algorithms to address truly rare events, so called 'black swans' [20], is extremely difficult and may not be practical. Third, while human factors are frequently cited as being sources for error in NDE applications, humans are inherently more flexible in handling unexpected scenarios and can be better at making such judgement calls. Human inspectors also have certain characteristics like common sense and moral values, which can be beneficial in choosing to the most reasonable and safest option. Humans in many cases can detect when an algorithm is making an extremely poor classification due to inadequate training and correct these errors.

Fourth, for many machine learning algorithms like DLNNs, it can be difficult to ascertain exactly why certain poor calls are made. These algorithms are often referred to as a 'black box', because the complex web of mathematical operations optimized for complex data interpretation problems do not generally lend themselves to back-engineering.

Approaches have be developed to sample the parameters space to ascertain the likely source for decisions [21]. As well, major research programs are currently pursuing the development of explainable artificial intelligence (XAI) [22]. Fifth, with the greater reliance on algorithms, it is quite possible that inspector skill will degrade over time. (This challenge of optimizing the human machine interface will be discussed in the next section.) Lastly, there is a clear potential for automated systems and algorithms to replace the work currently performed by inspectors; thus, there is a concern about unemployment and the impact to society with the greater use of algorithms for NDE decision making. Ideally, automation will enable greater productivity of the workforce, which should be a positive benefit in the long-run for society. However, the rapid transition of automation and how to manage these changes related to employment is complicated.

TABLE 1. The 'pros' and 'cons' of the application of algorithms in NDE.

Pros:	Cons:
1. Handle Laborious and Repetitive Tasks	1. High Cost (Development, Validation)
2. Error Reduction (Especially for Complex Tasks)	2. Cannot Make Decisions Well for Scenarios Not Trained (Poor Performance for Rare Events)
3. Faster Decisions / Follow-up Actions	3. Lack of Inherent Flexibility / Poor at Judgement Calls
4. Reduction in Overall Risk	4. Difficulty for Determine Reason for Poor Calls
5. Act as 'Digital Assistant'	5. Degradation of Human Skills
6. Repository for Human 'Expertise'	6. Unemployment

PERSPECTIVE ON INTELLIGENCE AUGMENTATION (IA) VERSUS ARTIFICIAL INTELLIGENCE (AI)

With recent progress and hype on the coming wave of artificial intelligence today, some perspective on AI is needed to understand how exactly these algorithms will be used by humans. While the vision for artificial intelligence has been to mimic human intelligence, in practice AI has only been successful for very focused tasks and has often required vast computing resources and power to achieve similar performance, relative to the human brain. While certain algorithms today can clearly perform better than humans for certain predefined and optimized tasks, it has not achieved the early goal of independent artificial intelligence. Humans not only have the capability to perform millions of different tasks, many in parallel run by the unconscious mind, but they also have the wherewithal to determine when it is appropriate to switch from one task to another and allow the consciousness mind to have awareness to specific thoughts and/or actions. The value of AI today is because humans are using it as a specific tool.

As a counterpoint to AI, Intelligence Augmentation (IA) refers to the effective use of information technology to enhance human intelligence [23-26]. This idea was proposed in the 1950s and 60s by early cybernetics and computer pioneers. During that same period, as the quest for full AI has encountered many fundamental obstacles, issues with IA seem moot. IA only uses technology as essentially 'support' for a human in to perform specific tasks. Relative to AI, IA has a long history of success. For example, consider the long history of information technology, from the birth of writing and slide rules, to smart phones and the internet today; all these forms of technology have essentially been developed to extend the information storage and processing capabilities of the human mind. Fundamentally, progress on AI algorithms should be viewed more as an evolution of tools to better support IA.

While most of the attention in recent years has been on the performance of AI over humans in games such as chess and recently Go [15], there are number of applications that have been cited where humans plus algorithms can exceed the performance of computer algorithms alone. Centaurs [25] and cyborgs [26] are terms used to refer to such IA collaborations. One example that is frequently cited is chess. A team of amateur chess players paired with three chess programs convincingly defeated a series a teams made up of chess grandmasters and some of the world's best chess programs [26,29]. While this case study is slightly dated and may not hold up to the success of AlphZero in the past year [28], fundamentally, all of these algorithms at some stage in their design for operational tasks have incorporated human input. This collaboration between humans and algorithms leveraging HPC has the potential to solve an array of greater problems than mere games of strategy. For example, the practice of engineering, for many decades, consists of humans leveraging their intellect with the support of computational tools to solve technical problems. Humans are still critical in asking the right questions and providing the appropriate focus, complementing the brute force computational power with creativity in selecting the most promising problem space to investigate. Humans also have a natural flexibility, versatility and intuition that AI systems have yet to achieve. These uniquely human qualities are still quite impressive, especially considering the relatively low power consumption of the human mind.

From the perspective of NDE applications incorporating algorithms, IA has the potential to address most of the disadvantages of AI cited above. For example, many of the most promising DLNN applications today, from speech recognition, to text translation and image classification, are still far from perfect. However, that does not mean that these tools are not useful. In practice, humans can frequently detect these errors by AI, and can quickly work around poor results. Humans often develop an understanding where such algorithms can be most appropriately applied, and where they should be avoided. By leveraging the algorithms where they are most useful, it becomes less critical for the algorithm to be able to handle all scenarios, especially very rare events. Lastly, by operators working in conjunction with algorithms, there is no need to pursue eliminating the human entirely. In general, the most cost effective and reliable solution will mostly likely be some form of hybrid, human plus machine, based approach.

However, there is a critical need to optimize the interaction between humans with computer algorithms in NDE. Some work has studied human machine interfaces (HMI) for different nondestructive evaluation applications [30-32]. For example, Bertović performed a detailed survey of prior work on human factors when interfacing with automation in NDT [31]. In this work, a Failure Modes and Effects Analysis (FMEA) was conducted to identify potential risks and preventive measures were proposed. Subsequent studies were used to verify the benefit of the preventive measures, highlighting mixed levels of success.

Additional guidance on the challenge of human machine interfacing can be gained from the experiences of other communities that also require high levels of reliability. For example, in aviation, the use auto-pilot systems and the transition between human control and auto-pilot is a pertinent case study for NDE. An example of the possible catastrophic consequences of automation is the tragic crash of flight AF447 [33]. In recent years, the accident rate for major jets was 1 major accident per 2.56 million flights. While overall air safety has been improving, incidents of the loss of control were not. Loss of control occurs when pilots fail to recognize and correct a potentially dangerous situation, causing an aircraft to enter an unstable condition. The most severe loss of control incident was the 2009 crash of Air France Flight 447, which killed 228 people [33]. Such incidents are typically triggered by unexpected, unusual events, often comprising multiple conditions that rarely occur together, that fall outside of the normal repertoire of pilot experience. Thus, this is the paradox of *almost totally safe systems*, where the same technology that allows systems to be efficient and largely error-free also creates systemic vulnerabilities that result in occasional catastrophes. Lessons learned from AF447 include (a) avoiding the cycle of implementing more automation to correct for poor human performance with existing auto-pilot systems (leading to poor human performance), (b) encouraging more hand-flying to prevent erosion of basic piloting skills, (c) improving the management of handovers from machines to humans, (d) increasing pilot training for rare events and (e) supplementing training using simulation of various rare event scenarios.

BEST PRACTICES FOR OPERATOR INTERACTION WITH ALGORITHMS

Building on this prior work and experience, a series of best practices for intelligence augmentation in NDE is proposed, highlighting how the operator should interface with NDE data and algorithms. Algorithms clearly have a great potential to help alleviate the burden of 'big data' in NDE; however, it is important that operators are appropriately involved in secondary indication review, and the detection of rare event conditions. Six best practices are proposed:

1. **Focus Initial Design of Algorithm(s) on Key NDE Indication Calls.** It is important to address the low-hanging fruit on implementing algorithms for NDE applications and to help alleviate the burden for inspectors of reviewing 'mostly good' data. The design of these algorithm requires a focus on the base capability for making NDE indication calls, to provide value and help ensure reliability. The algorithm design will consider necessary engineering development time, cost for acquiring necessary data, and the approach with the highest likelihood of success.
2. **Inspectors Provide Secondary Review of Indications and Review Data for Rare Events.** While there is an initial desire to have NDE algorithms make all indication calls and present simple 'green light' / 'red light' results; based on prior experience, additional information is always requested by engineering and management to understand the details on why an indication call was made. Inspectors need an interface to review each call with supporting data and provide feedback on the significance of the call relative to the technical requirements. As well, since no algorithm will be perfect, inspectors need to have a straightforward means to review the NDE data for 'black swan' events [20]. This entails identifying rare indications and determining when the acquisition of the NDE data is out of specification
3. **Provide Inspectors with Seamless Interface for Data Review, Reporting and Feedback**. Usability of human machine interfaces is a critical aspect of the NDE technique that must not be neglected. Ideally, inspectors need a way to provide feedback on the called indications using the same software interface for data review. Such data with operator feedback could potentially be used to refine the base algorithms in the future, so management of 'meta-

data' on missed and false calls is important. Frequently, there are means to annotate on the NDE image results and also include notes in some form of exported report.

4. Build Trust Over Time and Evaluate Cost-Benefit for Future Algorithm Enhancements. The transition of algorithms should initially be a phased approach, to both validate the algorithm performance and build an understanding of where the algorithms are reliable and where limitations exist. By tracking called indications over time, it becomes feasible to refine algorithm, as necessary. Cost-benefit analysis will play a role on this decision.

5. Software and Algorithms Can Support NDE Reliability through Process Controls. Simply demonstrating POD capability does not ensure reliability of the technique [34]. Failure Modes and Effects Analysis (FMEA) should be performed for all NDE techniques incorporating automation, to understand the potential sources for poor reliability [31]. In practice, NDE reliability depends on a reproduceable calibration procedure and a repeatable inspection process [34]. Process controls and algorithms can thus be used to ensure all calibration indications are verified and track key metrics that the NDE process is repeatable over time and under control.

6. Develop Software Interface to Support Operator Training ('NDE Simulator'). There is a potential to leverage the same software interface for training purposes, by having the operators periodically train and test their skills with various conditions in NDE data. Specific rare events can be stored and introduced periodically, as part of the regular re-training of inspectors. Thus, the interface could be used in a similar way that flight simulators are used for pilots to verify their performance under standard conditions and rare events.

Several case studies are presented in the following sections highlighting these best practices of leveraging algorithms and addressing human interfaces for several USAF NDE applications.

CASE STUDY 1 – 'AI' VISION BECOME 'IA' IN PRACTICE

Following the success of the C-141 weep hole inspection program [35], the development of automated data analysis algorithms were investigated for the inspection of beam cap holes in C-130 aircraft, as shown in Fig 2(a) [36]. Here, the fastener sites of interest are in locations of limited accessibility from the external surface and contain fasteners with sealant, as shown in Fig 2(b). Due to limitations with in NDE capability at the time, there was a need to develop improved ultrasonic techniques to detect fatigue cracks at these locations. A key challenge was the ability to discern multiple signals originating from a possible crack and a geometric feature in a part that are either closely spaced or superimposed in time. The C-130 beam cap holes provided a special challenge given the skewed riser, installed fasteners, and limited transducer accessibility of the B-scan inspection [see Fig 2(b)]. This inspection problem frequently produced reflections from the fastener hole occurring at similar times-of-flight as near and far crack signals. To address this challenge, a novel feature extraction methodology was developed to detect the relative shift of signals in time for adjacent transducer locations due to differing echo dynamics from cracks and part geometries [37]. This technique was the first ultrasonic NDE method using assisted analysis methods, validated through a POD study, to inspect for fatigue cracks on Air Force structures [36]. A view of the operator user interface is presented in Fig 2(c). While the vision for the approach was to have the automated data analysis (ADA) algorithms make all of the indication calls, during transition, some severe structure plus fastener conditions were found to produce false calls on rare occasions. To mitigate these rare false calls, the algorithm results and raw data required secondary review by inspectors. Although this technique was the first AI / neural-network based approach used to inspect a portion of the C-130 fleet, this case study is actually a very good example of intelligent augmentation in practice.

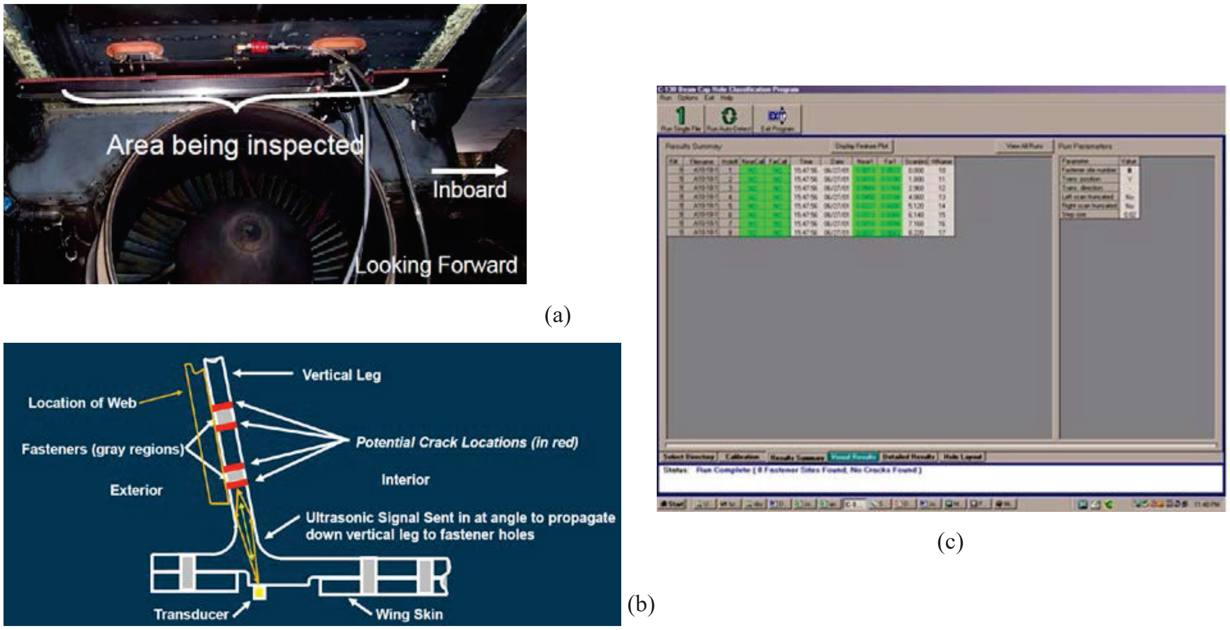

FIGURE 2. (a) Photo of C-130 beam-cap hole B-scan inspection, (b) diagram of inspection problem and (c) user interface for automated data analysis (ADA) software incorporating neural network classifiers [36].

CASE STUDY 2 – OPTIMIZING HUMAN MACHINE INTERFACE (HMI)

The ultrasonic inspection of aerospace polymer matrix composites is one of the most effective methods to detect critical defect types and ensure the reliability of composite structures. Most inspection applications of composites are based on pulse-echo ultrasonic testing and manual C-scan data interpretation. Using amplitude and time-of-flight C-scan data, delaminations, disbonds, porosity, and foreign materials can be detected and located in depth. However, the ultrasonic inspection of large composite structures requires significant manpower and production time. To address this inspection burden, automated data analysis software tools were developed and implemented [38]. The automated data analysis minimizes the inspector burden performing mundane tasks and allocate their time to analyze data of primary interest. When the algorithm either detects a feature in the data that is unexpected or that is found to be representative of a defect, then the indication is flagged for further analysis by the inspector.

A software interface for the automated data analysis (ADA) toolkit is shown in Fig. 3. The main view provides a summary of the found indications in the analyzed data, a visual presentation of an indication map, and quantitative metrics assisting the operator in understanding why each call was made. An example of ADA processing results is reported in the interface display shown in Figure 3. Options are provided to enter feedback into the ‘review’ column, to indicate if certain calls are incorrect. This example specimen contains artificial defects that have been added at varying locations and ply depth, including above and below the adhesive layer. Indications are listed in the spreadsheet display in the upper left and corresponding numbers are presented identifying the indications in the C-scan image display on the right. For these ADA evaluations for the two different scan orientations, the three triangular inserts in the bond region were all correctly called. The left most triangle is in front of the bond and the right two triangles are behind the bond. Indications for the six inserted materials at the radii are also observed in the TOF map in Figures 3. As well, there are options to add un-called indications, as missed calls, into the ADA report with comments.

Features are also provided to support verification of calibration scans. The first step is to evaluate the inspection system using calibration data acquired from a known reference standard. A signal-to-noise acceptance criterion is evaluated for the back-wall signal with respect to the internal noise echoes (e.g. 20 dB SNR) to ensure the transducers are operating properly. As well, call performance on known defects in the reference standard can be verified using the interface. Using filename indicators, an ADA verification process can be run to ensure all of the known inserts in the calibration standard that meet the call criteria are correctly identified by the ADA algorithm.

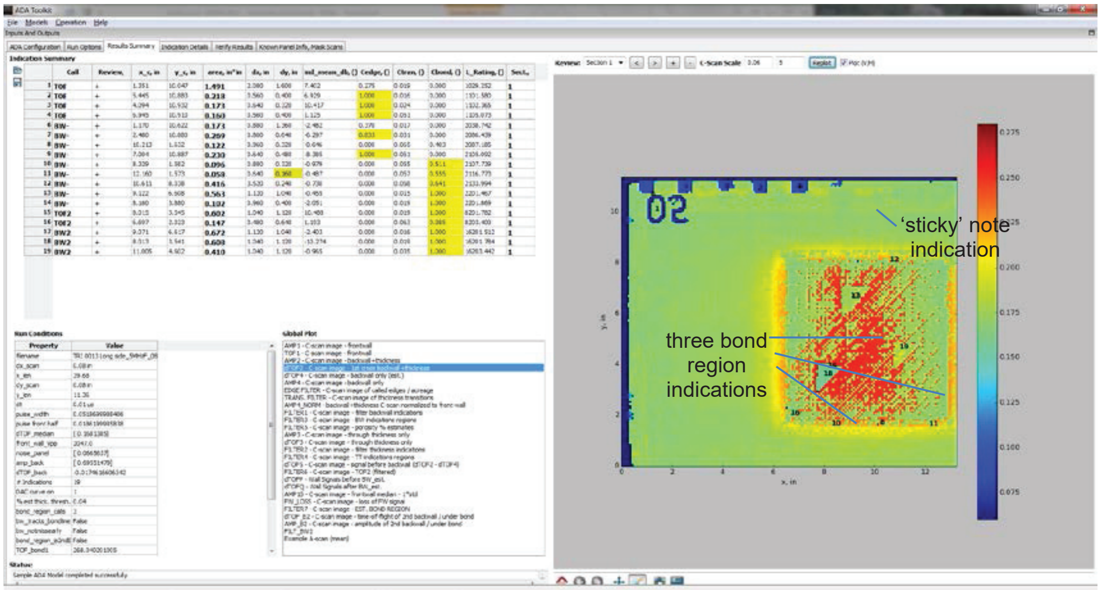

FIGURE 3. Example ADA toolkit interface results for a test panel, scanned from tool side, with TOF C-scan view [38].

CASE STUDY 3: ALGORITHM SUPPORT FOR IMPROVED PROCESS CONTROLS

The use of eddy current (EC) techniques to detect damage in aircraft structures and propulsion components is a key part of USAF programs to ensure that the risk of failures meets the desired requirements. There is a growing need to also accurately evaluate the dimensions of any detected cracks to properly guide repairs and support Digital Thread and Digital Twin programs. The current practice to size cracks in terms of surface length and depth uses the eddy current response amplitude along with an underlying assumption about the aspect ratio based on empirical data from POD specimens. This approach can be prone to error in the crack profile assumption and will be sensitive to varying part and test conditions. Recent progress has been made on model-based inverse methods with eddy current inspection for sizing surface-breaking cracks in propulsion components. However, variations in probe characteristics, such as coil orientation and offset resulting from probe liftoff or coil setback, can cause additional uncertainty in predicted dimensions. Recent work on model-based inverse methods with eddy current inspections of surface breaking discontinuities has shown increased sizing error due to variability in probes with the same design specifications [39]. To address sensitivity to probe variability, a model-based calibration process was introduced that estimates the state of the probe as a separate inverse problem. This step performs a six-parameter inversion with the experimental calibration data, addressing four probe state parameters (set-back/liftoff, tilt in two directions and rotation) with amplitude/phase estimation. This contrasts with past calibrations using only two parameters, for amplitude and phase. Using this process control for evaluation the eddy current probe state, the potential for improved crack inversion performance was demonstrated for certain varying probe tilt and liftoff conditions [39]. In practice, any probes with extreme winding and liftoff conditions can be detected and removed from production inspections.

DISCUSSION ON VALIDATION OF IA AND AI

NDE techniques, whether incorporating AI algorithms, manual inspector data review or a mixed IA based approach, require validation through a probability of detection (POD) evaluation. Comprehensive probability of detection (POD) evaluation procedures [40-41] have been developed to validate the reliability of NDI techniques, regardless of how the indication call is made. In prior work, a POD study was used to evaluation the capability of an automated data analysis (ADA) algorithms to detect cracks around holes in vertical riser aircraft structures [35]. In the study, the ADA approach incorporating neural networks was compared with manual data review by inspectors. Results demonstrate that the automated neural network approach was significantly improved in both detectability, false call rate and inspection time relative to manual data interpretation [35].

The greatest challenges with validation of NDE algorithms is ensuring that the algorithm is not over-trained, but can handle the variability for practical NDE measurement 'outside of the laboratory'. Testing algorithms with independent samples from training data is critical. Model-assisted approaches can also help to provide a diversity of conditions beyond what is practical and cost effective with experimental data only [40]. Because of these challenges, properly validating NDE techniques using IA is expected to be far easier to achieve than purely AI-based technique. Another example challenge problem is validating IA and AI self-driving car technology. For the time being, self-driving car technology simply augmenting the driver experience is more straightforward than fully validating a pure AI based self-driving car technology. Recent accidents in 2017-2018 during testing of self-driving cars indicate the care that is truly needed to properly validate such automated systems when people's lives are at stake.

CONCLUSIONS AND RECOMMENDATIONS

In recent years, advances have been made in the field of machine learning and artificial intelligence (AI), primarily through developments in deep learning neural networks (DLNN). The paper discusses the pros and cons of implementing algorithms for NDE applications. As a counterpoint to AI, intelligence augmentation (IA) was presented as an effective use of information technology to enhance human intelligence. Based on prior experience, this paper introduces a series of best practices for intelligence augmentation in NDE, highlighting how the operator should interface with NDE data and algorithms. Algorithms clearly have a great potential to help alleviate the burden of 'big data' in NDE; however, it is important that operators are involved in both secondary indication review, and the detection of rare event indications not addressed well by typical algorithms. Several past examples of transitioning algorithms for NDE applications are presented, emphasizing the successful interfacing of operator and software for optimal data review and decision making. Future work should continue to study the validation of NDE techniques that leverage both human and algorithms for data review and investigate appropriate process controls and software design to ensure optimum performance.

ACKNOWLEDGMENTS

Portions of this work were funded by the U.S. Air Force Research Laboratory, Materials and Manufacturing Directorate through the contract FA8650-15-D-5231 TO 001 with UTC.

REFERENCES

1. E. A. Lindgren, "US Air Force perspective on validated NDE – Past, present, and future," *42nd Annual Review of Progress in Quantitative Nondestructive Evaluation*, AIP Conf. Proc., **1706**, p. 020002 (2016).
2. R. J. Katt, ed., *Applying Materials State Awareness to Condition-based Maintenance and System Life Cycle Management*, National Academies Press, Washington, D.C., (2014).
3. J. D. Achenbach, "Quantitative nondestructive evaluation," *Int J of Solids and Structures*, **37**, n 1, pp. 13-27, (2000).
4. M. Schneider, R. Meyer, B. Nasralla, D. Piotrowski, "Panel: Automation for Sustainment - Emerging Technology and Opportunities," *Aircraft Airworthiness & Sustainment Conference,* Jacksonville, FL, (April 26, 2018).
5. P. Kobryn, et. al., "Digital thread and twin for systems engineering: EMD to disposal," *Proceedings of the 55th AIAA Aerospace Sciences Meeting*, Grapevine, TX, (January 2017).
6. E. A. Lindgren, "Opportunities for nondestructive evaluation: quantitative characterization," *Materials Evaluation*, **75**, n 7 p. 862 (2017).
7. C. W. Magnuson, D. Forsyth, "Digitizing the MRB Process for Composite Aerospace Structures," 45th Annual Review of Progress in QNDE, AIP Conference Proceedings, (to be published, 2019).
8. C. Buynak, E. Medina, and M. Jackson, "Digital NDI Data Capture and Reporting to Enhance Aircraft Sustainment," Aircraft *Airworthiness & Sustainment Conference,* Jacksonville, FL, (April 26, 2018).
9. D. Forsyth, J. C. Aldrin, C. W. Magnuson, "Turning Nondestructive Testing Data into Useful Information," Aircraft *Airworthiness & Sustainment Conference,* Jacksonville, FL, (April 26, 2018).
10. I. Jahanzaib and J. Jasperneite. "Scalability of OPC-UA down to the chip level enables "Internet of Things"," In *Industrial Informatics (INDIN), 2013 11th IEEE International Conference on*, pp. 500-505, (2013).
11. N. G. Meyendorf, L. J. Bond, J. Curtis-Beard, S. Heilmann, S. Pal, R. Schallert, H. Scholz and C. Wunderlich, "NDE 4.0—NDE for the 21st century—the internet of things and cyber physical systems will revolutionize NDE", 15th *Asia Pacific Conference for Non-Destructive Testing*, (2017).
12. R Link and N. Riesss, " NDT 4.0 - Significance and Implications to NDT – Automated Magnetic Particle Testing as an Example," *12th European Conference on Non-Destructive Testing (ECNDT 2018), Gothenburg 2018, June 11-15*, web site: https://www.ndt.net/search/docs.php3?showForm=off&id=22997.
13. J. C. Aldrin and E. A. Lindgren, "The Need and Approach for Characterization - US Air Force Perspectives on Materials State Awareness," *44th Annual Review of Progress in QNDE*, AIP Conference Proceedings, **1949**, p. 020004 (2018).
14. K. Fukushima, and S. Miyake. "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition." *Competition and Cooperation in Neural Nets*, Springer, Berlin, Heidelberg, pp. 267-285, (1982).

15. G. Lewis-Kraus, "The great AI awakening," *The New York Times Magazine,* (December 14, 2016).
16. G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, **18**, n 7, pp. 1527-1554 (2006).
17. Y. LeCun, B. Yoshua, and G. Hinton. "Deep learning," *Nature,* **521**, n 7553, pp. 436-444, (2015).
18. A. Stapleton, "DARPA's Latest Take on AI Illuminates the Course of Intelligent Assistance," (March 1, 2017), web site: https://opusresearch.net/wordpress/2017/03/01/darpas-latest-take-on-ai-illuminates-the-course-of-intelligent-assistance/.
19. M. Yao, "Understanding the limits of deep learning," web site: https://venturebeat.com/2017/04/02/understanding-the-limits-of-deep-learning, (April 2, 2017).
20. N. N. Taleb, "Black swans and the domains of statistics." *The American Statistician*, **61**, n 3, pp. 198-200 (2007).
21. J. D. Olden, and D. A. Jackson. "Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks," *Ecological modelling*, **154**, n1-2, pp. 135-150, (2002).
22. D. Gunning, "Explainable artificial intelligence (XAI)," *Defense Advanced Research Projects Agency (DARPA) I2O, Program Update,* (November, 2017).
23. P. Skagestad, "Thinking with machines: Intelligence augmentation, evolutionary epistemology, and semiotic," *Journal of Social and Evolutionary Systems*, **16,** n 2, pp. 157-180 (1993).
24. A. Rastogi, "Artificial Intelligence - Human Augmentation is What's Here and Now," *Medium*, (Jan 12, 2017), web site: https://medium.com/reflections-by-ngp/artificial-intelligence-human-augmentation-is-whats-here-and-now-c5286978ace0.
25. N. Case, "How to become a centaur" *Journal of Design and Science*, Issue #3-CASE, (2018).
26. M. Kitces, "What cyborg chess can teach us about the future of financial planning," (July 24, 2017), web site: https://www.kitces.com/blog/cyborg-chess-advisor-teach-about-future-financial-planning/.
27. S. Makridakis, "The forthcoming artificial intelligence (AI) revolution: Its impact on society and firms." *Futures*, **90**, pp. 46-60 (2017).
28. P. Gerbert, "AI and the 'Augmentation' Fallacy," *MIT Sloan Management Review*, (May 16, 2018), web site: https://sloanreview.mit.edu/article/ai-and-the-augmentation-fallacy/.
29. T., Cowen, *Average Is Over: Powering America Beyond the Age of the Great Stagnation*, Penguin, (2013).
30. D. D. Dudenhoeffer, D.E. Holcomb, B.P. Hallbert, R. T. Wood, L. J. Bond, D. W. Miller, J. M. O'hara, E. L. Quinn, H. E. Garcia, S. A. Arndt, and J. Naser, *Technology Roadmap on Instrumentation, Control, and Human-Machine Interface to Support DOE Advanced Nuclear Energy Programs*. INL/EXT-06-11862. (2007).
31. M. Bertović, "Human factors in non-destructive testing (NDT): risks and challenges of mechanised NDT," PhD. Dissertation, (2016). Tu Berlin, Germany.
32. N. G. Meyendorf, R. Schallert, S. Pal, and L. J. Bond. "Using remote NDE, including external experts in the inspection process, to enhance reliability and address today's NDE challenges," *7th European-American Workshop on Reliability of NDE*, (2017).
33. N. Oliver, T. Calvard, and K. Potočnik. "The tragic crash of Flight AF447 shows the unlikely but catastrophic consequences of automation," *Harvard Business Review,* (2017), Web site: https://hbr.org/2017/09/the-tragic-crash-of-flight-af447-shows-the-unlikely-but-catastrophic-consequences-of-automation.
34. W. D. Rummel, "Nondestructive inspection reliability—history, status and future path," *Proceedings of the 18th World Conference on Nondestructive*, Durban, South Africa, pp. 16-20, (2010).
35. J. C. Aldrin, J. D. Achenbach, G. Andrew, C. P'an, B. Grills, R. T. Mullis, F. Spencer, and M. Golis, "Case Study for the Implementation of an Automated Ultrasonic Technique to Detect Fatigue Cracks in Aircraft Weep Holes," *Materials Evaluation*, **59**, n 11, p.1313-1319, (2001).
36. E. A. Lindgren, J. R. Mandeville, M. J. Concordia, T. J. MacInnis, J. J. Abel, J. C. Aldrin, F. Spencer, D. B. Fritz, P. Christiansen, R. T. Mullis, and R. Waldbusser, "Probability of Detection Results and Deployment of the Inspection of the Vertical Leg of the C-130 Center Wing Beam/Spar Cap", 8th Joint DoD/FAA/NASA Conference on Aging Aircraft, (2005).
37. J. C. Aldrin, C. V. Kropas-Hughes, J. Knopp, J. R. Mandeville, D. Judd, E. Lindgren, "Advanced echo-dynamic measures for the characterisation of multiple ultrasonic signals in aircraft structures", *Insight*, **48**, n 3, pp. 144-148, (2006).
38. J. C. Aldrin, D. S. Forsyth, J. T. Welter, "Design and Demonstration of Automated Data Analysis Algorithms for Ultrasonic Inspection of Complex Composite Panels with Bonds," *42nd Annual Review of Progress in Quantitative Nondestructive Evaluation*, AIP Conf. Proc., **1706**, p. 020002 (2016).
39. J. C. Aldrin, E. B. Shell, E. K. Oneida, H. A. Sabbagh, E. Sabbagh, and R. K. Murphy, S. Mazdiyasni, E. A. Lindgren, R. D. Mooers, "Model-based probe state estimation and crack inverse methods addressing eddy current probe variability," *43rd Annual Review of Progress in QNDE,* AIP Conf. Proc., **1806**, p. 110013, (2017).
40. US Department of Defense, "Nondestructive evaluation system reliability assessment," MIL-HDBK-1823A, (2009).
41. ASTM E3023-15 *Standard Practice for Probability of Detection Analysis for â Versus a Data*, (2015).
42. J. C., Aldrin, C., Annis, H. A.,Sabbagh, and E. A., Lindgren, "Best Practices for Evaluating the Capability of Nondestructive Evaluation (NDE) and Structural Health Monitoring (SHM) Techniques for Damage Characterization," *42nd Annual Review of Progress in QNDE, Incorporating the 6th European-American Workshop on Reliability of NDE*, **1706**, p. 200002, AIP Publishing, (2016).