Flaw Detection in Ultrasonic Data Using Deep Learning

Virkkunen, Iikka; Koskinen, Tuomas

Published in:
Baltica XI

Published: 01/01/2019

Document Version
Publisher's final version

Flaw Detection in Ultrasonic Data Using Deep Learning

Iikka Virkkunen[1] & Tuomas Koskinen[2]

[1]Aalto University
Finland

[2]VTT Technical Research Centre of Finland Ltd
Finland

Abstract

Non-destructive testing often relies heavily on expert judgement to detect flaws in noisy data. This is the case also for ultrasonic testing, where small flaws can be detected by seasoned NDT inspectors even when the signal to noise ratio is small. While automating this kind of flaw detection has long been tried, it has not seen widespread success yet.

In recent years, machine learning models have proven successful in various image pattern recognition and classification tasks. In particular, deep learning models and convolutional neural networks (CNN's) have proven capable of tackling tasks long considered infeasible for computers. The task of defect detection in ultrasonic signal is not unlike these recently solved tasks and many of the advances in related fields help also in defect detection.

In this paper, we present results from training of deep CNNs using mechanized ultrasonic data augmented with virtual flaws (eFlaws). Virtual flaws provide a way to augment ultrasonic data in far more realistic ways than previously done before and we show, that modern deep convolutional network can be trained successfully to detect flaws using virtual flaw data. The performance of the trained model is compared with human performance (POD on similar data). The trained models show superhuman performance with a clear margin and thus show great promise for wider application.

1. Introduction

Non-destructive evaluation (NDE) is constantly pushed to its limits. The smaller are the flaws, that can be detected, the more time is available for mitigation. Alternatively, the inspections can be made less frequent and thus more cost-efficient. At the same time, the reliability requirements of the inspections are quite high. Often, probability of detection (POD) exceeding 90% need to be demonstrated at 95% confidence level. Thus, consistency in performance is also very important.

Automated systems have long been used for various NDE systems. These provide consistent results and do not show the variation commonly seen in human inspectors due to fatigue, stress or other factors. However, the traditional automated systems have relied on simple decision algorithms such as a signal amplitude threshold. In more demanding inspection cases, such as the typical in-service inspections, the human inspectors achieve far superior inspection results than such simplified automated systems. Consequently, most of these inspections are currently analyzed by human experts, even when the data acquisition is highly automated. Such analysis is time consuming to do and taxing for the personnel.

The key problem with more sophisticated automation has been, that the work of the human inspector does not lend itself to simple algorithmic description. The inspectors acquire their skill through years of training and utilize various signal characteristics in their judgement (e.g. the “signal dynamics”). Machine learning systems have now been used for a long time to automate systems, where simple algorithmic description is intractable. For these systems, instead of hand-coding the decision algorithm, it is “learned” through iterative adjustment of a highly flexible decision function, such as a neural network. The field is well established, and well described in textbooks ([1-3]) basic building blocks of the currently dominant machine learning algorithms have been used for decades, the improvements in computational power (GPU acceleration, in particular) have enabled more complex and powerful models that reach near human-level performance in tasks like image classification, machine translation and the like.

Early attempts to use machine learning for NDT flaw detection and classification focused on using simple neural networks to classify various types of NDT data. Masnata and Sunser [4] used a neural network with single hidden layer to classify various flaw types (cracks, slag inclusinos, porosity) from ultrasonic A-scans. Before learning, the A-scan was reduced to 24 pre-selected features using the Fischer discriminant analysis. Chen and Lee [5] used wavelet decomposition, to obtain features from A-scans and reported potential, while the training and testing was done with limited data set. Yi and Yun [6] similarly used shallow neural network to train flaw type classifier with a larger data set. Although in many cases this early work reported high classification accuracy, the results proved to be difficult to scale and to extend to new cases.

With the increase in computational power, the used machine learning models have become more powerful. Many authors have reported good results with shallow models like support vector machines (SVM’s). While these models offer high classification capability, they also require a pre-selected set of features to be extracted from the raw NDT signal. Fei et al. [7] used wavelet packet decomposition of ultrasonic A-scans to train SVM for defect classification in petroleum pipeline. Sambath et al. [8] used neural network with two hidden layers to classify ultrasonic A-scans using a hand-engineered set of 12 features. Shipway et al. [9] used random forests to detect cracks from fluorescent penetrant inspections (FPI). Cruz et al. [10] used feature extraction based on principal component analysis to train a shallow neural network to detect cracks from ultrasonic A-scans.

One of the issues with developing ML-models for defect classification has been the limited availability of training data. Liu et al. [11] used finite element simulation results to provided artificial NDT signals to augment training data. Munir et al. [12] used deep CNN’s to classify austenitic stainless steel welds. The training data was obtained from weld training samples containing artificial flaws (i.e. solidification flaws). The data-set was augmented by shifting the A-scans in time-domain and by introducing Gaussian noise to the signal.

The primary benefit of the deep models is, that they are able to learn abstract feature representation during training and do not require hand-tuned or pre-selected features to be extracted from the signal before the training. Recently Meng et al. [13], Zhu et al. [14] and Munir et al. [12,15] used deep convolutional neural networks (CNNs) for defect classification in ultrasonic and EC-data. Meng et al. [13] used deep neural networks with an SVM top layer for enhanced classification capability. The classifier was used to classify voids and delamination flaws in carbon fiber composite material. Before presented to the CNN, the raw A-scan data was decomposed using wavelet packet decomposition and the resulting coefficients re-organized into 32x16 feature matrix. Thus, the CNNs classified the A-scans separately.

Zhu et al. [14] used deep CNN’s to detect cracks in eddy current signal. Also, drop-out layer was used to estimate the confidence of the classification, which is an important opportunity in using ML in field NDT, where the reliability requirements are very high. This work is also notable in that the raw signal database was exceptionally representative with NDT indications representing plant data for various defect types [16].

In summary, the current state of the art for using machine learning in NDT classification may be seen to focus on two distinct aims. Firstly, modern shallow ML models (e.g. random forests) with advanced feature-engineering are used with the aim to develop computationally lightweight models that can be implemented on-line to aid inspector in manual inspection. Secondly, very deep CNNs are used to learn from raw NDT signals without the need for explicit feature engineering. The recent work on deep models takes full advantage of recent advances in models developed for other industries and shows good results across different NDT fields. For ultrasonic testing, the existing machine learning models have mostly involved classification in single A-scan level.

In the present work, we used virtual flaws to implement inspection-aware data augmentation. The use of virtual flaws enables generation of highly representative augmented data set for ML applications. The virtual flaws have also been used for human training and thus there is good confidence, that their representativeness is high.

Currently, most demanding inspections are done by human inspectors. Thus, it is of interest to compare the machine learning models with human inspector performance. However, in many cases even the human inspection performance is not quantified and known with sufficient reliability to allow direct comparison to developed ML models. In present work, we used human performance data obtained from previous research [17] and developed the machine learning models to work on comparable data thus enabling direct comparison between human inspector and modern machine learning model.

2. Experimental

For this study, ultrasonic data from previous research was used. The data was previously used to develop simplified POD-tool for inspector training. Thus, the performance of human inspectors is known and documented for these data [17].

2.1 Ultrasonic data

Inspected specimen for data-acquisition was a butt-weld in austenitic pipe. Three thermal fatigue cracks with depths 1.6, 4.0 and 8.6 were implemented in the inner diameter of the pipe near the weld root by Trueflaw ltd. and scanned with ultrasonic equipment. Inspection method used for data acquisition was Transmission Receive Shear (TRS) phased array, one of the common methods used in inspecting of austenitic and dissimilar metal welds. The scan was carried out by using Zetec Dynaray 64/64PR-Lite flaw detector linked to a PC. The probes used were Imasonic 1.5 MHz 1.5M5x3E17.5-9 matrix probes with central frequency at 1.8 MHz, element dimensions 3.35 x 2.85 mm and element arrangement as 5 x 3 elements. The wedge was ADUX577A to produce a shear wave more efficiently and only one linear scan without any skew angles were utilized. The ultrasonic wave was focused to the inner surface of the pipe and the probe was positioned in a way that the beam would be focused directly to the manufactured cracks. Coupling was applied through a feed water system and the pipe was rotated underneath the probe to assure constant and even coupling between the probe and the pipe. Probe position was carefully monitored along the scan line by Zetec pipe scanner with 0.21 mm scan resolution. The specimen and the inspection procedure is described in more detail in Koskinen et al. [18]. The specimen and the scanner can be seen in Figure 1. Due to data efficiency, the only the angle where the cracks were the most visible was chosen for further analysis. This, in this case was the 45° scan.

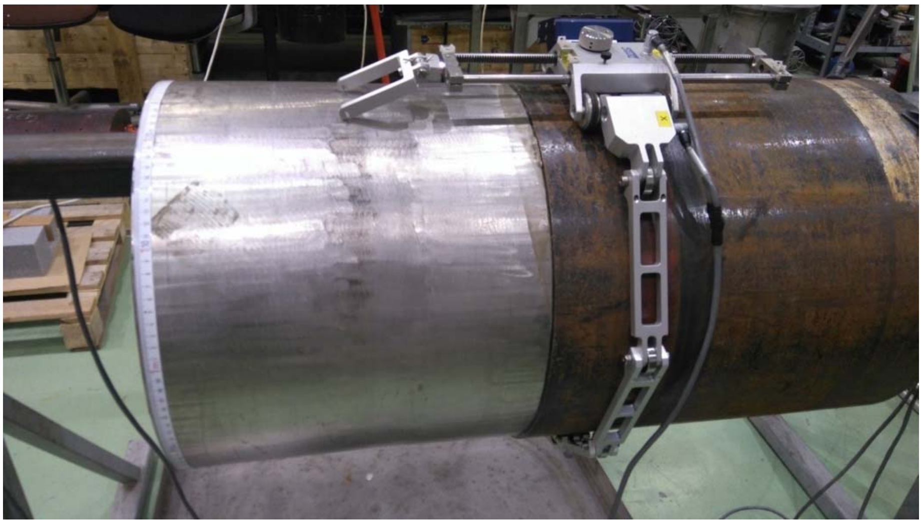

Figure 1. The used ultrasonic scanner set-up. The test block is seen on the left side. Steel extender was used to support the scanner during scanning due to limited length of the test block.

2.2 Data augmentation using virtual flaws

Virtual flaw augmentation was used to broaden the representative sizes of the cracks. The virtual flaw software used was Trueflaw's eFlaw. Details of the eFlaw technology are explained in more detail in [18-21]The teaching data set was created the similar way as for testing data set for human inspectors in previous paper [17].

The single 45°scan line data containing signals from three manufactured thermal fatigue flaws was taken as the source data for training the machine learning model. This is the same data, that was used to generate human POD results in [17]. From this data, large number of data files were generated using the same algorithm as previously. The data contained 454 A- scans each containing 5058 samples with 16 bit depth.

For machine learning purposes, the data was further processed. Each A-scan was cut so that only the interesting area around the weld was included resulting in 454 x 454 point data. Then, the resolution of the ultrasonic data was down-sampled to 256 x 256 points.

Altogether 20000 variations were generated to be used as training and validation data. The data was stored in minibatches of 100 UT-images per file with accompanying true state information showing the included crack state (whether there is included an introduced virtual flaw and the effective size of the flaw). The data set also contained data, where virtual flaw process had been used to copy unflawed section to another location. This was done to avoid and to detect the possibility that the machine learning model would somehow learn to notice the virtual flaw introduction process, instead of the actual flaws.

Typical images of training data with flaws are shown in Figure 2.

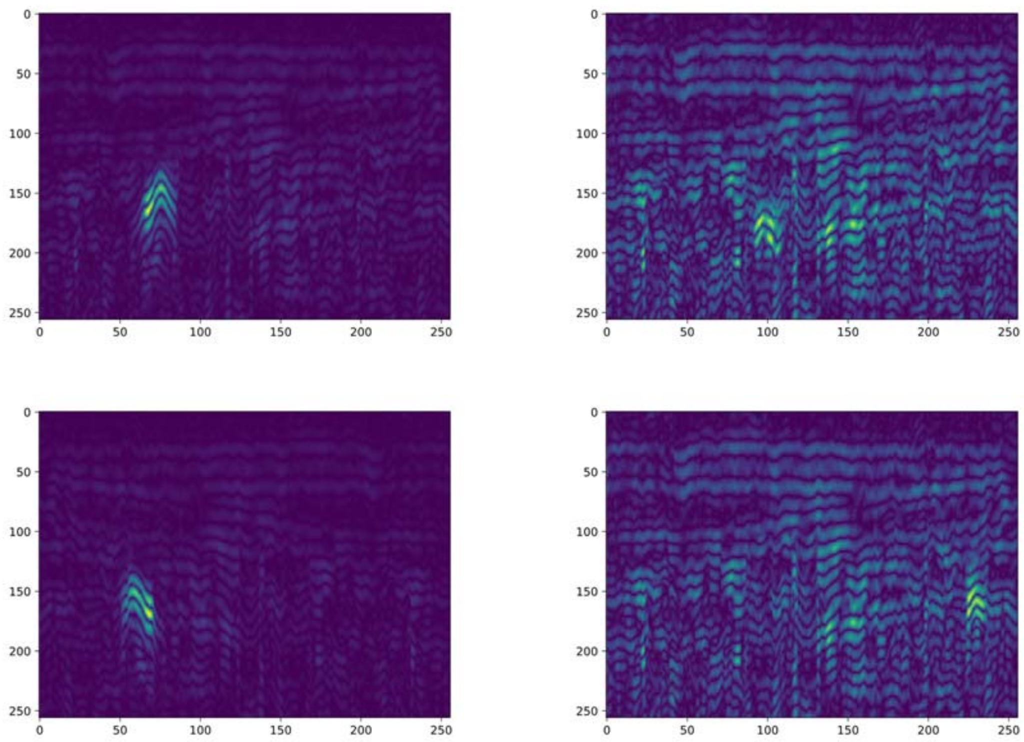

Figure 2. Example training images with cracks introduced using eFlaw.

2.3 Machine learning classifier

The machine learning architecture used is based on the VGG16 network [22]. For ultrasonic data analysis, the basic network was augmented with a first max-pooling layer, with pooling size adjusted to the wavelength of the ultrasonic signal. This max-pooling layer had the effect of removing spectral information from the image so that the rest of the network was left with an envelope amplitude curve. The training used binary cross entropy as the cost function and training was done using the RMSProp [23]. The computation was implemented with the Keras library [24] using TensorFlow back-end [25]. The network is illustrated in Figure 3.

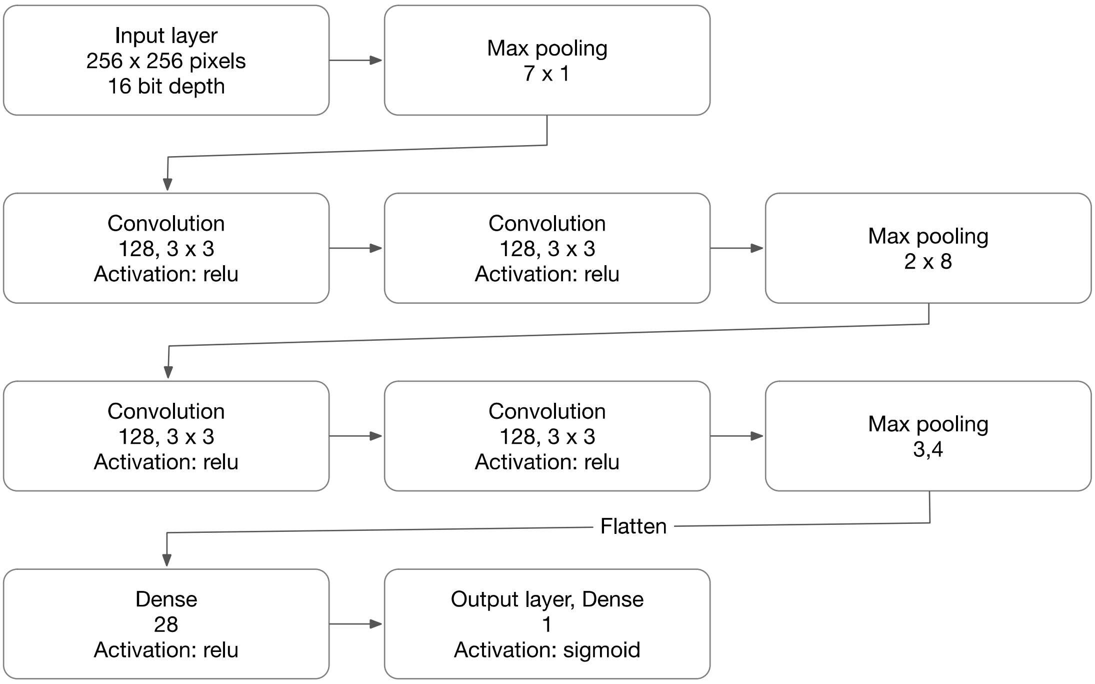

Figure 3. Used machine learning classifier.

3. Results

The network was trained for 100 epochs of 10000 samples. This resulted in perfect classification: all cracks were correctly classified and no false calls were made. The number of training epochs was set by hand to stop slightly after perfect classification score was achieved. During development, the results were evaluated against a separate validation set. The final result was then evaluated against a previously unseen verification set. Each set contained 100 images, with roughly 50% cracks.

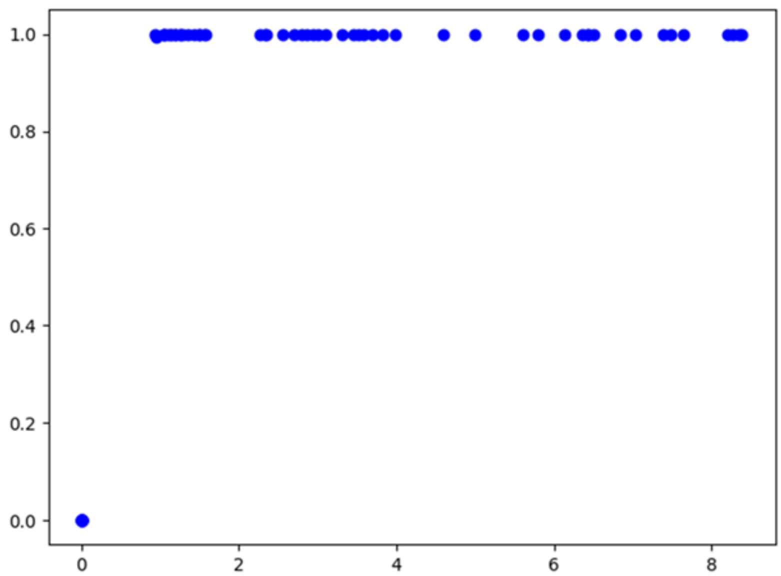

Figure 4. Classifier results. All the cracks were correctly classified with wide variation in effective size. (The effective size was computed as in [17]). No false calls were made.

4. Discussion

To evaluate the network performance against human performance, data set from previous work was utilized [17]. The performance was evaluated using MIL-HDBK-1823a hit/miss analysis [26].For the machine learning classifier, no misses were made, so the MIL-HDBK-1823a analysis can not be used directly. For the ML classifier, the smallest found crack (i.e. the smallest crack in the set) was used as reference performance measure. In practice, the $a_{90/95}$ values are somewhat lower than the largest missed flaw, and thus this metric is conservative. For human inspectors, the $a_{90/95}$ was used as the performance metric. The performance comparison is summarized in table 1.

Table 1. Comparison of human performance (from[17]) and current ML performance.

	Detectable crack size	**False calls**
Human inspectors	1–2.5	<30
ML-classifier	0.9	0

The results show that the current deep convolutional networks are powerful enough to achieve flaw detection surpassing human inspectors, when trained with the same data. However, it should be noted, that the number of real flaws in the original UT data is very small and not sufficient to represent natural variation in cracks. Likewise, the amount of unflawed data is limited. Consequently, the trained network should not be used as-is for real inspections. Instead, these results provide evidence, that if provided with sufficient training data, such modern networks are able to achieve or exceed human-level performance. Therefore, modern machine learning models offer huge opportunity for UT inspections and further research needs to be conducted.

5. Conclusions

The following conclusions can be drawn from this study:

- Deep convolutional neural networks are powerful enough to reach and exceed human-level performance in detecting cracks from ultrasonic data
- Data augmentation using virtual flaws is seen as key enabling technique to train machine learning networks with limited flawed data

6. Acknowledgements

The data augmentation using virtual flaws and the initial network & training was contributed by Trueflaw Ltd. Their contribution is gratefully acknowledged.

References

1. Andrew Gelman, J.B.C., Hal S. Stern, David B. Dunson, Aki Vehtari, Donald B. Rubin, *Bayesian Data Analysis*. 3 ed. Chapman & Hall/CRC Texts in Statistical Science. 2013: Chapman and Hall/CRC.

2. Chollet, F., *Deep Learning with Python*. 2017: Manning Publications Co. 363.

3. Goodfellow, I.B., Yoshua; Courville, Aaron, *Deep Learning* Adaptive Computation and Machine Learning series. 2016, Cambridge, MA: MIT Press. 778.

4. Masnata, A. and M. Sunser, *Neural Network Classification of Flaws Detected by Ultrasonic Means.* NDT&E International, 1996. **29**(2): p. 87-93.

5. Lee, C.H.C.G.G., *Neural networks for ultrasonic nde signal classification using time-frequency analysis.* IEEE, 1993.

6. Yi, W. and I.-s. Yun, *The Defect Detection and Non-Destructive Evaluation in Weld Zone of Austenitic Stainless Steel 304 Using Neural Network-Ultrasonic Wave.* KSMME International Journal, 1998. **12**(6): p. 1150 - 1161.

7. Fei, C., Z. Han, and J. Dong, *An ultrasonic flaw-classification system with wavelet-packet decomposition, a mutative scale chaotic genetic algorithm, and a support vector machine and its application to petroleum-transporting pipelines.* Russian Journal of Nondestructive Testing, 2006. **42**(3): p. 190-197.

8. Sambath, S., P. Nagaraj, and N. Selvakumar, *Automatic Defect Classification in Ultrasonic NDT Using Artificial Intelligence.* Journal of Nondestructive Evaluation, 2010. **30**(1): p. 20-28.

9. Shipway, N.J., et al., *Automated defect detection for Fluorescent Penetrant Inspection using Random Forest.* NDT & E International, 2019. **101**: p. 113-123.

10. Cruz, F.C., et al., *Efficient feature selection for neural network based detection of flaws in steel welded joints using ultrasound testing.* Ultrasonics, 2017. **73**: p. 1-8.

11. Liu, S.W., et al., *Detection of Cracks using Neural Network and Computational Mechanics.* Computer methods in applied mechanics and engineering, 2002. **191**: p. 2831 - 2845.

12. Munir, N., et al., *Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments.* Journal of Mechanical Science and Technology, 2018. **32**(7): p. 3073-3080.

13. Meng, M., et al., *Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks.* Neurocomputing, 2017. **257**: p. 128-135.

14. Zhu, P., et al., *A novel machine learning model for eddy current testing with uncertainty.* NDT & E International, 2019. **101**: p. 104-112.

15. Munir, N., et al., *Convolutional neural network for ultrasonic weldment flaw classification in noisy conditions.* Ultrasonics, 2018.

16. Udpa, L.R., P., *Steam Generator Management Program: Automated Analysis of Array Probe Eddy Current Data*. 2009, EPRI, Palo Alto, CA.

17. Virkkunen, I., et al., *Effect of Feedback and Variation on Inspection Reliability*. 2017, German Society for Non-Destructive Testing.

18. Koskinen, T., et al., *Producing a POD curve with emulated signal response data.* Insight, 2018. **60**(1): p. 42-48.

19. Svahn, P.-H., et al., *The use of virtual flaws to increase flexibility of qualification*. 2018, NDT.net.

20. Virkkunen, M., et al. *Feasibility Study of Using eFlaws on Qualification of Nuclear Spent Fuel Disposal Canister Inspection*. 2016.

21. Virkkunen, I., K. Miettinen, and T. Packalén. *Virtual flaws for NDE training and qualification*. 2014.

22. Simonyan, K. and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition.* eprint arXiv:1409.1556, 2014: p. arXiv:1409.1556.

23. Zeiler, M.D., *ADADELTA: An Adaptive Learning Rate Method.* eprint arXiv:1212.5701, 2012: p. arXiv:1212.5701.

24. Chollet, F. *Keras*. 2015; Available from: https://keras.io.

25. Martín Abadi, A.A., Paul Barham, Eugene Brevdo,, et al. *TensorFlow: Large-scale machine learning on heterogeneous systems*. 2015; Available from: https://tensorflow.org.

26. Annis, C., *Nondestructive Evaluation System Reliability Assessment*, in *Department of Defense Handbook*. 2009. p. 171.