Proceedings of the ASME 2023
50th Annual Review of Progress in
Quantitative Nondestructive Evaluation
QNDE2023
July 24-27, Austin, Texas

QNDE2023-XXXXX

COMPARATIVE STUDY ON DEEP LEARNING METHODS FOR DEFECT IDENTIFICATION AND CLASSIFICATION IN COMPOSITE AEROSTRUCTURE MATERIAL

Austin Yunker*
Argonne National Laboratory
Lemont, IL

Rami Lake
Northern Illinois University
Dekalb, IL

Rajkumar Kettimuthu
Argonne National Laboratory
Lemont, IL

Zach Kral
Spirit AeroSystems
Wichita, KS

ABSTRACT

Aircraft structures are required to have a high level of quality to satisfy their need for light weight, efficient flight, and withstanding high loads over their lifespan. These aerostructures are typically made from composite material due to their good tensile strength and resistance to compression. To ensure their structural integrity, the composite material requires inspection for common flaws such as porosity, delaminations, voids, foreign object debris, and other defects. Ultrasonic testing (UT) is a popular non-destructive inspection (NDI) technique used for effectively evaluating composite material. Current inspection methods rely heavily on human experience and are extremely time consuming. Therefore, there is a need for the development of techniques to reduce the manual inspection time. This work compares the performance of different deep learning-based methods in the identification and classification of defects. Deep learning has shown great promise in numerous fields, and we show its effectiveness in the evaluation of composite aerostructure material. The methods developed here are both highly reliable with a top recall value of 98.63% as well as extremely efficient requiring an average of 4 seconds during the inferencing stage to evaluate new composites.

Keywords: Non-destructive insepction (NDI), ultrasonic testing (UT), deep learning, composites, defects

1. INTRODUCTION

State-of-the-art aircraft are designed using fiber-reinforced composites due to their good tensile strength and resistance to compression [1]. Specifically, carbon-fiber-reinforced plastic (CFRP) is applied to wing planks, fuselages, and sandwich panel skins [2]. These materials are continuously improving in product efficiency, cost-effectiveness, and are commonly used in applications to load-bearing structures outside the aerospace domain including wind turbines, transportation, medical equipment, and so on. Manufacturing of composite materials is a multivariable task, involving many procedures, where various types of defects may occur within a composite product, giving rise to significant safety concerns in service. Therefore, it is critical to evaluate the structural health of the composite material ensuring the safety of the manufactured piece [3]. Defects resulting in a loss of mechanical properties can occur in all composite structures and tend to increase in frequency with structural complexity. These defects include voids, porosity, inclusions, and delamination. Defects can arise both at production and during service [3]. Robust and reliable non-destructive testing (NDT) of composites is essential in the production stage to catch and repair defects. There are numerous NDT techniques built upon different principles. Ultrasonic (UT) testing is one such technique. UT testing is suitable for detecting and sizing flaws that are embedded within the surface by generating signal waveforms used for inspection. Other advantages include its ability to propagate through thick solid parts, detection of various flaws, and various testing standards are in place, ensuring its trusted use for critical parts [4].

The evaluation of these ultrasonics is typically done through trained inspectors which can be both time consuming and prone to human error. After a fuselage section is scanned, evaluation is done visually through software developed to convert the ultrasonics data into meaningful visualizations. Generally, there are three ways in which to view the scanned section: A-scans, B-scans, and C-scans. A-scans are plots of amplitude versus time and can be collected to generate B-scans and C-scans. B-scans are 2D images where the color scale represents amplitude, and two axes represent two dimensions – typically horizontal distance moved

*Corresponding author: ayunker@anl.gov

along the component and vertical distance (depth) into the part. A C-scan is also a 2D image, but it represents a 3D volume with the two axes representing two horizontal dimensions while the color scale represents a particular amplitude value [4]. The C-scan is the default in evaluating scanned sections. In the evaluation process, fuselage regions are verified with the use of composite standards, controlled representative structures with known defect placements.

Advanced computer vision techniques, particularly those based on machine learning algorithms, can provide new perspectives on the high-level visual understanding of universal tasks. The power of these techniques suggests a new approach to evaluating UT images, in which machine learning algorithms rather than manual inspection is used for the general detection and classification of manufacturing imperfections. An advantage of using machine learning systems is their ability to take highly dynamic and non-linear data spread across numerous features and find the relationship between inputs (waveform of ultrasound response) and desired output (classification of defects). Deep learning, a subfield of machine learning devoted to the study of neural network algorithms, has shown near human level performance on numerous classification tasks [5, 6]. Applied to the field of NDT, an early work [7] built three artificial neural networks (ANNs) for location, quantification, and classification of structural damage using vibrational characteristics in aluminum cantilever beams. More recent work [8] has provided a comparison of various computer vision techniques, showing the superior performance of deep learning models compared to conventional methods when applied to stainless steel plates for defect identification. N. Saeed, et al. [9] implemented a convolutional neural network (CNN) to detect artificially created sub-surface defects in a CFRP sample as well as a deep feed-forward neural network to estimate the defect depth using thermograms. The work in [10] develop a CNN to detect flaws from phased-array ultrasonic data while also making extensive use of data augmentation to enhance the limited data by utilizing virtual flaws. Finally, Y. Duan, et al. [11] apply a long short-term memory recurrent neural network (LSTM-RNN) to classify defects occurring in honeycomb materials using a thermography-based NDT technique.

Artificial intelligence (AI) has seen widespread use in numerous fields and industries and continues to be used in new ways. Previous work described above shows that non-destructive evaluation of composite material appears to be one such field that can be improved with the integration of AI methods. This work provides a comparative study on the use of different deep learning architectures to classify defects in composite standard structures modelled after state-of-the-art aircrafts. The remainder of this study is organized as follows. Section 2 introduces the necessary materials and methods used. In Section 3, the corresponding results are reported, and methods are compared and discussed. In Section 4, concluding remarks are made with possible directions for future work given.

2. MATERIALS AND METHODS

2.1 Ultrasonics

Non-destructive testing (NDT) covers a wide range of analytical techniques to inspect, test or evaluate chemical/physical properties of a material, component or system without causing damage. Early established NDT techniques include ultrasonic, X-ray radiography, liquid penetrant testing (LPT), magnetic particle testing and eddycurrent testing, which were initially developed for steel industry. Among these, ultrasonic and radiographic detection are also effective inspection techniques for composite structures [3]. It is difficult to select appropriate NDT techniques for a specific purpose; however, ASTM E25335 serves as a practical guide in using NDT methods on composite materials/structures for aerospace applications [12].

To date, there have been numerous NDT methods based on different principles. They can be categorised into five groups: (1) VI (i.e. those visible to the human eye); (2) acoustic wave–based techniques, such as AE, nonlinear acoustics and ultrasonic waves; (3) optical techniques, which include IRT, THz testing, shearography, DIC; (4) imaging-based techniques, for example, X-ray/neutron radiography/tomography and micro-tomography; (5) electromagnetic field–based techniques, such as eddy-current testing, remote field testing, magnetic particle inspection and magnetic flux leakage testing [3].

Here, our focus is on ultrasonic testing (UT). UT is an acoustic inspection technique that operates through surface wave testing, bulk wave testing and guided wave propagation with guided wave analysis being superior for anisotropic material. When used for NDT inspection of composite material, UT uses normal incidence longitudinal waves and is particularly sensitive to in-plane defects. It operates in three detection modes, reflection, backscattering, and transmission of pulsed elastic waves in a material system. It introduces guided high-frequency sound waves, ranging from 1 kHz to 30 MHz to effectively detect flaw size, crack location, delamination location, fiber waviness, meso-scale ply fiber orientation and layup stacking sequence. A typical UT system consists of a transmitter and receiver circuit and transducer tool and display device, see Fig 1. Ultrasonic arrays are widely used in place of single element transducers in many NDE applications in which single element transducers are combined. The advantage being both speed and the ability to generate B-scans [3].

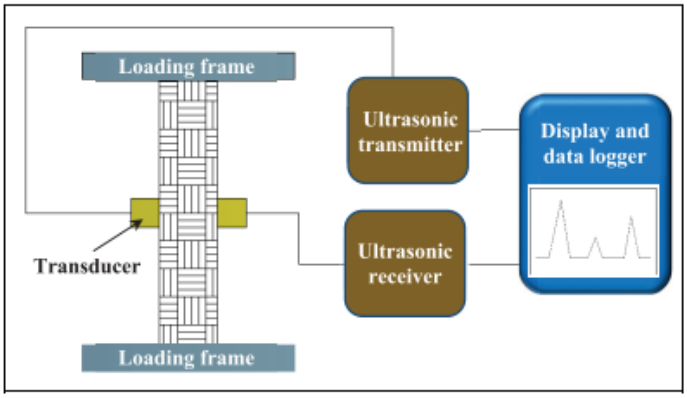


FIGURE 1: GENERAL ULTRASONIC TESTING OF A COMPOSITE MATERIAL

2.2 NDT Data

As mentioned above, inspectors evaluate ultrasonics data visually using A-scans, B-scans, and C-scans. C-scans are used by default as they produce a view of the entire material with the pixel color representing a particular amplitude value. The software used for converting ultrasonics data into C-scans automatically generates different C-scan views of the same material based on thresholding and gating different amplitude values. This allows inspectors to quickly evaluate the material by considering different views of the same region. For example, a given threshold scheme can be used to concentrate on defects that may occur at the top of the surface while others can be used to examine the inner portion of the composite.

To generate these automatic views, controlled representative structures known as standards are used, similar to phantoms used in the medical domain. These standards are artificially created composite pieces with hand-placed defects at a variety of locations and depths with the goal of enveloping any real defect in the actual structural part. Figure 2 shows the C-scan results for the same standard but with two different threshold schemes applied. The standard seen in Fig. 2 has seven regions of different ply counts with similar damage placement, rectangular and square regions, containing three different materials to represent damage categories. The colors assigned are based on an adjustable data range representing signal changes recorded by the sensor. Here, the colors increase in gray scale based on increased signal changes within the given data range. Signal changes above the data range are assigned red. By adjusting the range, different regions of the standard are highlighted as possible regions of interest, explaining the difference in color assignments of the same standard in Fig. 2.

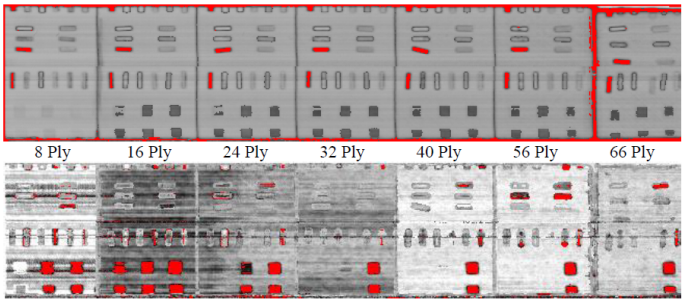
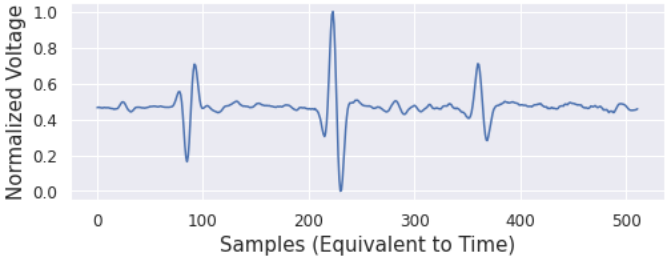
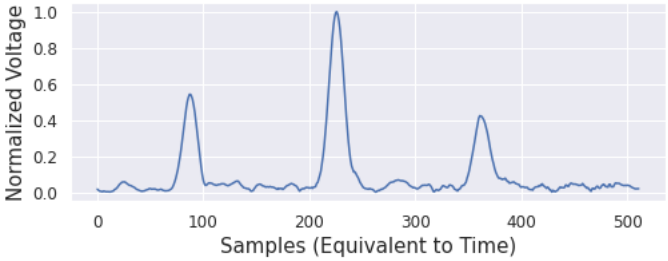


FIGURE 2: C-SCAN OF STANDARD WITH MULTIPLE THRESHOLD SCHEMES

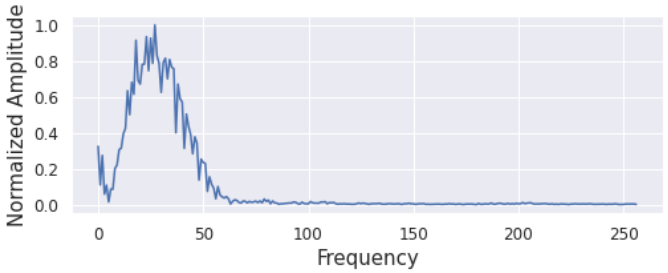
Each scanning point, equivalent to a pixel, of the C-scan has an associated waveform of 512 samples that contains the measured amplitude of ultrasonic response stored as a 16-bit value. The standards have 128x648 scanning points and are saved as a 3D volume with dimensions 128x648x512. While the original signal can be used as input into a model, we also consider the performance of the deep learning models under different data pre-processing techniques. We consider the original raw signal, the analytical signal generated using the Hilbert transformation, and finally the transformation of the signal into the frequency domain using the discrete Fourier Transform (DFT) as well as the combination of the previous two. The analytical signal produces a signal with less variation which may increase model performance. By converting the time series signal into its frequency components, we are able to model the response in a different domain. Figure 3 shows an example of an input signal into the model. Figure 3a represents the original unaltered signal while the Fig. 3b represents the analytical signal and Fig. 3c representing the transformation of the signal into the frequency domain using the DFT.



(a) Raw Signal



(b) Analytical Signal



(c) Discrete Fourier Transform

FIGURE 3: MODEL INPUT

2.3 Deep Learning Architectures Under Evaluation

Due to their artificial nature, each generated standard has a corresponding ground truth label that allows for easy training, evaluation, and comparison of different deep learning architectures. When viewed as a C-scan, we can formulate the problem as a semantic segmentation task in which the goal is to assign a label to each pixel, with each sample (512) of the waveform considered as a separate channel of an image (natural images use three channels for red, green, and blue). A common deep learning architecture used for semantic segmentation is the U-Net, built on 2D convolutional layers, which is used in this study. Since each pixel contains sufficient information to classify defects, we can reformulate the problem as a time series classification problem in which each A-scan is treated as an independent sample. The deep learning architectures considered in this reformulation

include the long short-term memory recurrent neural network (LSTM-RNN), convolutional neural network based on 1D convolutional layers, temporal convolutional network (TCN), and the recent transformer network. Finally, each A-scan can be transformed from the time series domain into the frequency domain using the discrete Fourier Transform where each frequency component is an independent feature. Here, we consider a simple multi-layer perceptron (MLP) network. By passing the models' final output through a SoftMax layer, we are able to obtain a vector of class probabilities with the highest being assigned as the class prediction. Each deep learning architecture considered is briefly discussed below.

2.3.1 U-Net. Originally developed for biomedical image segmentation, the U-Net architecture consists of a down-sampling network followed by an up-sampling network [13]. The down-sampling network is comprised of 2D convolution layers followed by an activation function, typically ReLU or LeakyReLU, ending with a pooling layer. The up-sampling network follows the same structure but switches convolution layers for transposed convolution layers. These convolutional layers work by sliding a kernel across an image that learns to extract a type of feature from its input. Finally, high-resolution features from the down-sampling path are concatenated to the up-sampled output [13]. The exact architecture used consists of three down-sampling/up-sampling layers with zero padding allowing the image sizes to be the same. The first down-sampling layer contains 32 channels while the last contains 128 with the network using the LeakyReLU activation function.

2.3.2 Long Short-Term Memeory Recurrent Neural Network. Specifically developed to model sequential input, recurrent networks (RNNs) can use their feedback connections to store representations of recent input events making them effective at modelling time series data. However, a major drawback of the RNN architecture is its inability to learn long-term relationships [6]. Long short-term memory RNNs (LSTM-RNN) networks were developed to overcome this. The LSTM-RNN architecture includes both a short-term and long-term state. The long-term state is controlled through the forget gate, input gate, and output gate that controls which parts should be erased, which parts should be added, and which parts should be read, respectively [6]. The architecture considered in this study consists of 3 LSTM layers with 25 neurons in each layer using the hyperbolic tangent activation function with the features being passed to a MLP with a single hidden layer.

2.3.3 1D Convolutional Neural Network. Like the U-Net, this architecture is built using convolutional layers. However, since the input is a 1D timeseries, 1D convolutional layers are used instead of 2D. Again, kernels are slid across the sequence learning to extract features from its input with a stride of two, effectively halving the sequence through each layer. A down-sampling path is used here as well but only consists of convolutional layers and activation functions, with no pooling layers. After the features are extracted, they are flattened and passed to a simple MLP. The specific architecture used here consists of four convolutional layers with the first containing 8 channels and the last having 32. Each layer uses the ReLU activation function which is finally flattened and passed to a MLP consisting of a single hidden layer.

2.3.4 Temporal Convolutional Neural Network. The temporal convolutional network (TCN) extends the convolutional neural network above by increasing its ability to model longer sequences, similar to LSTM extending RNN. This is done using dilated convolutions. Simple convolutions are only able to look back at a sequence with size linear in the depth of the network. To overcome this, TCNs employ dilated convolutions that enable exponentially large receptive fields allowing the network to model much longer sequences [14]. The TCN used here consists of 5 hidden layers with 25 channels each which is then flattened and passed into a MLP with one hidden layer.

2.3.5 Transformer. Introduced in the context of natural language processing, transformers no longer make use of recurrent or convolutional layers and are instead composed solely of self-attention mechanisms. These self-attention mechanisms allow relating different positions of a single sequence to compute a representation of the sequence. Using more than one self-attention mechanism gives a multi-head attention. Finally, a positional encoding is introduced to inject information about the relative position of the tokens in the sequence [15]. The specific architecture used includes 5 multi-attention heads compressing the original sequence before passing it into a MLP with a single hidden layer.

2.3.6 Multi-Layer Perceptron. The last architecture under consideration is the multi-layer perceptron (MLP). The MLP consists of simple fully connected linear layers that map the input into hidden layers before finally giving an output [16]. In between these linear layers are activation functions and dropout layers that allow the model to learn non-linear relationships while not overfitting the data. The linear layers assume no relationship among the input unlike recurrent and convolutional layers modeled specifically for time series and images, respectively. However, the input for this architecture will be the frequency components after applying the discrete Fourier Transform making this approach appropriate. The MLP considered consists of 10 hidden layers with ReLU activation and a dropout value of 0.3.

2.4 Performance Evaulation Criteria

In order to compare the different deep learning architectures, we use commonly reported classification metrics. By comparing the predicted results with ground truth labels, we can define four major statistics:

- True Positive (TP): number of defect signals correctly detected.
- True Negative (TN): number of normal signals classified as non-defect.
- False Positive (FP): number of normal signals incorrectly classified as defect.
- False Negative (FN): number of defect signals incorrectly classified as normal.

From these statistics, we can define four metrics, Accuracy (Acc), Precision (Pr), Recall (Re), and F1-score (F1):

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Pr = \frac{TP}{TP + FP} \tag{2}$$

$$Re = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = 2 \cdot \frac{Pr \cdot Re}{Pr + Re} \tag{4}$$

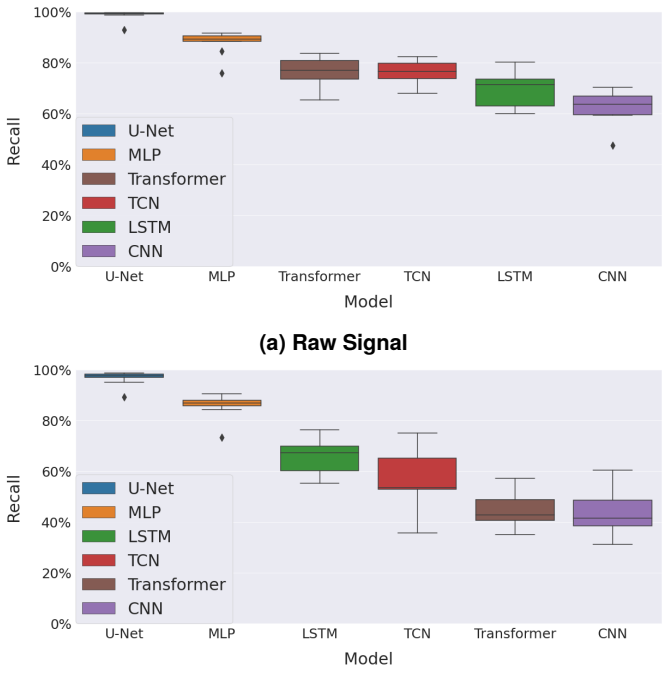
Accuracy measures the overall correctness of the model. Precision measures the ratio of actual defects to the number of predicted defects while recall measures the ratio of correctly identified defects. The F1-score is the harmonic mean of the precision and recall. In the context of defect detection, recall is typically the most important as most applications cannot risk any defects being missed ensuring a reliable system. Therefore, final models will be ranked by their recall value. Lastly, the wall-time required for training and inferencing each model will be reported.

3. RESULTS AND DISCUSSION

To generate the results, we used 11 standards for training and validation and 9 standards for testing. The goal with this setup is to simulate a possible future direction of training, validating, and testing on three separate aircrafts. The input shape for the vision-based U-Net is (11, 512, 128, 648) representing the number of standards, input channels, height, and width respectively. Since the time series-based methods treat each signal independently, we can flatten the first, third, and fourth dimension above giving an input shape of (912384, 512) representing the number of signals and number of samples within each signal. When transforming into the frequency domain, the input shape becomes (912384, 257) representing the number of samples and the number of frequency components within each sample. The training of each model was done using a single NVIDIA A100 GPU card with the number of epochs set to 1000 and a learning rate of 0.0005. The batch size for the U-Net was set at 2 so that at each update step 2 standards were used during training. For the time series models, the batch size was set at 2048. It was found that a smaller batch size led to poor convergence due to the class imbalance in which non-defect signals made up a majority of the dataset. Finally, all models were developed using the PyTorch framework, trained with the Adam optimizer and finally saved based on the lowest validation error [17, 18].

Table 1 shows the full results considered in this study grouped by the data pre-processing method (model input) ranked in descending order by recall. The results are averaged over the 9 standards used for testing. Overall, the results show that all deep learning models considered can identify defects with accuracies ranging from 90% - 100%. However, there is a higher degree of variation in the recall metric ranging from 44%-99%. Furthermore, the difference in recall between the top two different models is 11%. Finally, 4a and 4b show the model results when using the raw and analytical signal, respectively.

We can also examine model performance by examining the model's output compared to the corresponding ground truth label



(a) Raw Signal

(b) Analytical Signal

FIGURE 4: RECALL BY MODEL

at the C-scan level. Figure 5 shows each model's prediction at the C-Scan level. Here, we see that each model is capable of identifying the general defect area but differe in the number of false positives as well as the boundary between the defect and non-defect region.

Table 2 shows both the wall-time for training and inference ranked in descending order by recall for each deep learning model considered with the inference time averaged over the 9 standards used for testing. While training time can be critical for applications requiring frequent retraining, we assume here that training time is negligible as models can be trained offline or in the background. Here, we are concerned with inference time, as models used in a production setting are expected to be both accurate and fast. Slow inference times make even the most accurate models impractical and counterproductive. We measure inference time in an end-to-end fashion starting from loading the data in to saving final model outputs. In Table 2, we see that most models report an average inference time of 3-6 seconds showing that each approach could be used in a production setting. Furthermore, the best performing model also has one of the fastest training times, which may be useful for certain applications. Other notable remarks based on the results include:

- Using the discrete Fourier Transform gives better results than both the raw and analytical signal,
- Vision-based approaches outperform time series-based approaches,
- Within the time series approach, the raw signal is preferred over the analytical signal,
- The LSTM, based on recurrent layers, is the slowest model

TABLE 1: TEST METRICS GROUPED BY DATA PRE-PROCESSING RANKED IN DESCENDING ORDER BY RECALL

Model Input	Model	Accuracy	Precision	Recall	F1-Score
Raw Signal					
DFT	U-Net	99.63%	98.92%	98.63%	98.77%
DFT	MLP	97.15%	91.41%	87.75%	89.53%
Raw Signal					
	U-Net	99.31%	98.18%	97.03%	97.59%
	Transformer	94.16%	81.13%	76.50%	78.61%
	TCN	95.50%	90.68%	76.03%	82.66%
	LSTM	94.07%	86.54%	69.67%	77.04%
	CNN	91.26%	71.91%	62.76%	66.84%
Analytical Signal					
DFT	U-Net	99.23%	97.99%	96.62%	97.29%
DFT	MLP	96.66%	89.97%	85.66%	87.76%
Analytical Signal					
	U-Net	99.28%	97.98%	96.80%	97.39%
	LSTM	93.42%	84.36%	66.16%	73.98%
	TCN	93.19%	91.05%	57.11%	69.35%
	Transformer	91.11%	85.82%	44.75%	58.47%
	CNN	89.99%	75.56%	43.81%	54.96%

to train and,

- The U-Net model has the lowest variation in recall as seen in Fig. 4.

TABLE 2: TRAINING AND INFERENCE TIME GROUPED BY DATA PRE-PROCESSING RANKED IN DESCENDING ORDER BY RECALL

Model Input	Model	Training[1]	Inference[2]
Raw Signal			
DFT	U-Net	22	4.79
DFT	MLP	77	3.65
Raw Signal			
	U-Net	30	4.16
	Transformer	155	3.27
	TCN	506	5.22
	LSTM	594	4.81
	CNN	89	4.69
Analytical Signal			
DFT	U-Net	26	5.98
DFT	MLP	78	4.74
Analytical Signal			
	U-Net	31	5.39
	LSTM	589	6.17
	TCN	506	6.30
	Transformer	148	4.43
	CNN	88	5.79

[1] Minutes, [2] Seconds

Overall, we've seen the effective use of different deep-learning based approaches in detecting defects in composite material. However, there appears to be a substantial benefit in formulating this problem using a vision based approach. The vision approach allows for taking advantage of the spatial correlation that exists between defects i.e., defects occur in clusters of pixels. The time series based approaches treat pixels independently and make predictions based on the signal alone. Therefore, they fail to include the information in the surrounding area. Finally, it is worthwhile to note that the defects used here are simple polygon regions (squares and rectangles) and are easy for a vision based approach to detect whereas actual defects may be of any shape.

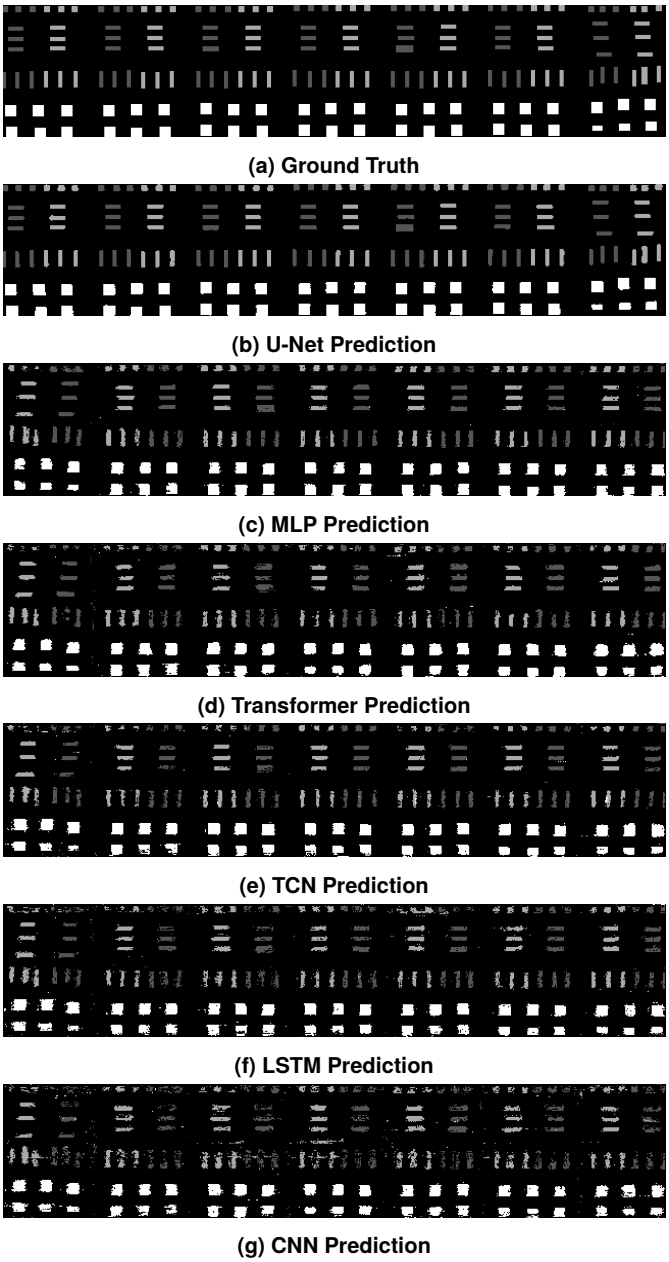
4. CONCLUSION

The purpose of this study was to evaluate and compare the performance of different deep learning models in identifying and classifying defects in composite material modeled after state-of-the-art aircrafts. The models considered include the: Multi-Layer Perceptron, Transformer, 1D Convolutional Neural Network, Temporal Convolutional Neural Network, Long Short-Term Memory Recurrent Neural Network, and U-Net. We also compared the performance of these models using different data pre-processing techniques including the analytical signal based on the Hilbert transformation as well as the discrete Fourier Transform. Our work shows both the effectiveness of different deep learning methods in defect identification and classification as well as the efficiency of each approach, requiring only a few seconds during the inferencing stage. Conventional NDI methods rely heavily on human experience; thus, they are subjective, do not scale, and are extremely time consuming. Therefore, there is a strong need for the integration of computer-based methods with deep learning being one such method. Future work seeks to extend the methods considered here to actual aircraft structures in production.

5. ACKNOWLEDGEMENTS

This work was supported in part by the U.S. Department of Energy, Office of Science, under contract DE-AC02- 06CH11357.

6. LICENSE



(a) Ground Truth

(b) U-Net Prediction

(c) MLP Prediction

(d) Transformer Prediction

(e) TCN Prediction

(f) LSTM Prediction

(g) CNN Prediction

FIGURE 5: C-SCAN PREDICTIONS USING RAW SIGNAL

gonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable world-wide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan. http://energy.gov/downloads/doe-public-access-plan.

REFERENCES

[1] Giurgiutiu, Victor. *Introduction*. Academic Press, London (2016): .

[2] Ibrahim, M.E. "Nondestructive Evaluation of Thick-Section Composites and Sandwich Structures: A Review." *Applied Science and Manufacturing* Vol. 64 (2014): pp. 36–48. URL https://doi.org/10.1016/j.compositesa.2014.04.010.

[3] Bing Wang, Tung-Lik Lee Kevin S Fancey, Shuncong Zhong and Mi, Jiawei. "Non-Destructive Testing and Evaluation of Composite Materials/Structures: A State-of-the-Art Review." *Advances in Mechanical Engineering* Vol. 12 No. 4. DOI 10.1177/1687814020913761. URL https://doi.org/10.1177/1687814020913761.

[4] Felice, Maria V. and Fan, Zheng. "Sizing of Flaws Using Ultrasonic Bulk Wave Testing: A Review." *Ultrasonics* Vol. 88 (2018): pp. 26–42. URL https://doi.org/10.1016/j.ultras.2018.03.003.

[5] Zhang, G.P. "Neural networks for classification: a survey." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* Vol. 30 No. 4 (2000): pp. 451–462. DOI 10.1109/5326.897072.

[6] Ian Goodfellow, Aaron Courville, Yoshua Bengio. *Deep Learning*. The MIT Press, London (2017).

[7] Carlos M. Ferregut, Roberto A. Osegueda and Ortiz, Jamie. "Artificial Neural Networks for Structural Damage Detection and Classification." *SPIE Proceedings* DOI https://doi.org/10.1117/12.207718.

[8] Jiaxing Ye, Shunya Ito and Toyama, Nobuyuki. "Computerized Ultrasonic Imaging Inspection: From Shallow to Deep Learning." *Sensors* Vol. 18 No. 11 (2018): p. 3820. DOI https://doi.org/10.3390/s18113820.

[9] Numan Saeed, Zafar Said, Nelson King and Omar, Mohammed A. "Automatic defects detection in CFRP thermograms, using convolutional neural networks and transfer learning." *Infrared Physics Technology* Vol. 102 (2019): p. 103048. DOI https://doi.org/10.1016/j.infrared.2019.103048.

[10] Ikka Virkkunen, Oskari Jessen-Juhler, Tuomas Koskinen and Rinta-aho, Jari. "Augmented Ultrasonic Data for Machine Learning." *Journal of Nondestructive Evaluation* Vol. 40 No. 1. DOI https://doi.org/10.1007/s10921-020-00739-5.

[11] Caiqi Hu, Shicai Liu-Yiqian Yan Ning Tao Ahmad Osman Clemente Ibarra-Castanedo Stefano Sfarra Dapeng Chen, Yuxia Duan and Zhang, Cunlin. "LSTM-RNN-based defect classification in honeycomb structures using infrared thermography." *Infrared Physics Technology* Vol. 102 (2019): p. 103032. DOI https://doi.org/10.1016/j.infrared.2019.103032.

[12] "Standard Guide for Nondestructive Examination of Polymer Matrix Composites Used in Aerospace Applications." ASTM E2553 (2017).

[13] Olaf Ronneberger, Philipp Fischer and Brox, Thomas. "U-Net: Convolutional Networks for Biomedical Image Segmentation." *ArXiv e-prints* (2015). URL https://arxiv.org/abs/1505.04597.

[14] Shaojie Bai, J. Zico Kolter and Koltun, Vladlen. "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling." *ArXiv e-prints* (2018). URL https://arxiv.org/abs/1803.01271.

[15] Ashish Vaswani, Niki Parmar-Jakob Uszkoreit Llion Jones Aidan N. Gomez Lukasz Kaiser, Noam Shazeer and Polosukhin, Illia. "Attention Is All You Need." *ArXiv e-prints* (2017). URL https://arxiv.org/abs/1706.03762.

[16] McCulloch, Warren and Pitts, Walter. "A Logical Calculus of the Ideas Immanent in Nervous Activity." *Bulletin of Mathematical Biophysics* Vol. 5 (1943): pp. 115–133. DOI https://doi.org/10.1007/BF02478259.

[17] Paszke, Adam, Gross, Sam, Massa, Francisco, Lerer, Adam, Bradbury, James, Chanan, Gregory, Killeen, Trevor, Lin, Zeming, Gimelshein, Natalia, Antiga, Luca, Desmaison, Alban, Köpf, Andreas, Yang, Edward, DeVito, Zach, Raison, Martin, Tejani, Alykhan, Chilamkurthy, Sasank, Steiner, Benoit, Fang, Lu, Bai, Junjie and Chintala, Soumith. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. Curran Associates Inc., Red Hook, NY, USA (2019): .

[18] Kingma, Diederik P. and Ba, Jimmy. "Adam: A method for stochastic optimization." *ArXiv e-prints* (2014). URL https://arxiv.org/abs/1412.6980.

[19] AeroSystems, Spirit. "Programs: Company." Spirit AeroSystems (2023). Accessed April 9, 2023, URL https://www.spiritaero.com/company/programs/.