DATA DRIVEN COMPOSITES: THE CHALLENGE AND PATHS FORWARDS

Muthu Ram Prabhu Elenchezhian, Vamsee Vadlamudi, Rassel Raihan, Kenneth Reifsnider*

Institute for Predictive Performance and Methodologies,
The University of Texas at Arlington Research Institute, Fort Worth, TX
*Corresponding Author: kenneth.reifsnider@uta.edu

ABSTRACT

The concept of “data driven” grew out of the general subject of data analytics, especially in the business world. The classical sequence proposed by Dykes (2010) shown below illustrates the general contents of the concept.

Indeed, the idea of “predictive modeling” of the future behavior of a system has strong roots in the business community. In general, for engineering and other fields, the data driven challenge is to use the past with information from the present to predict the future behavior of a system. In turn, the user must pay for collection and processing, hosting and maintenance of the data, and for the cost of analysis, etc. and address the risk of breach of the system.

The first and perhaps the most important question in this venture is the nature of the data itself, i.e., the “source of truth.” We have many measurable quantities for composite materials; which of those should we use for the objectives of our analysis? Traditional data such as “failure rates” result in sparse data sets that may be impossible to analyze, and abundant data from “health monitoring” systems may be distantly related to the physics of our objective function. Data interpretation and analysis for composites is also a challenge; there may be “missing physics” that motivates the use of machine learning or more general artificial intelligence / neural network systems for interpretation. The present paper discusses these general questions and discusses several paths forward, including the implementation of mixed physics / neural network analysis approaches, machine learning, material state variable definitions for composite materials, and recent experience with these and other concepts.

1. INTRODUCTION

The typical life cycle of composite materials evolves from design of the layup and the structure, analysis on the design to calculate the stresses and strains by classical theories or finite element models, manufacture of the composite part by current state of art manufacturing
techniques, inspection of the quality through nondestructive technology, and testing of the structure for further development or use of the part with established structural health monitoring techniques. With the increased use of the composites in the aerospace, automobile, construction and material industries, the need for increasing the speed of these processes, i.e. to reduce the time, cost, and risk taken for each process by advanced methodologies increases. As part of a solution to these requirements, the Data Driven approach, sometimes called the Artificial Intelligence approach of the modern era, serves as a best fit. In this research work, we discuss the basic methodologies of artificial intelligence approaches, and the current state of art work done on each part of this life cycle. We also discuss the issues faced by developing such data driven composites and also the future of these, and the concept of obtaining “missing physics” from such data driven approaches. Finally, we include some examples of data types and methods of recovery that serve this general purpose. Figure 1 describes the different stages in life cycle of the composites, in which the data driven methodologies can play a vital role.

Figure 1: Lifecycle of Composites

2. ARTIFICIAL INTELLIGENCE

2.1 Machine Learning

In the context of machine learning, the most important elements of the method are the prediction algorithms, which give the predictions of the defining probabilities based on the trained data. The basic algorithms used in the machine learning are the regression and classification algorithms. As general regression modeling deals with the linear equations, using models such as neural networks, and polynomial input features, gives the ability to derive complex non-linear equations.
2.1.1 Regression

Linear regression is widely used in statistics, data analysis, machine learning, biological and industrial engineering to develop possible relationships when the physics of the data is unknown.

2.1.1.1 Hypothesis

For a given input vector $X$, we write the predicted output of the linear regression as

$$h(x) = W \ast X + b$$  \[1\]

where $W$ is the transpose of weight vector and $b$ is a scalar term, also called as bias unit.

2.1.1.2 Cost Function

For linear regression, we calculate the cost function from the loss function (also called the error function) as the mean of the squared errors.

For the $i^{th}$ example, the Loss Function is given by

$$L(h(x), y) = \frac{1}{2} (h(x^i) - y^i)^2$$  \[2\]

Hence the Cost Function is given by $J$ where

$$J(W, b) = \frac{1}{2m} \sum_{i=1}^{m} (h(x^i) - y^i)^2 + \frac{\lambda}{2m} \sum_{j=1}^{n} W_j^2$$  \[3\]

and $h(x^i)$ is the predicted output and $y^i$ is the actual output, and $\lambda$ is the regularization parameter.

2.1.2 Classification

2.1.2.1 Hypothesis

For the given input vector $X$, we write the predicted output of the logistic regression as

$$\hat{y} = g(W \ast X + b)$$  \[4\]

where $W$ and $b$ are the parameters or weights.

2.1.2.2 Cost function

The aim of the cost function is to minimize error and approximate the prediction output $\hat{y}$ to the actual output $y$. Hence, to obtain this, the Loss function is defined as

$$L(\hat{y}, y) = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y}))$$  \[5\]

The cost function $J$ is given by
\[ J(W, b) = \frac{1}{m} \sum_{i=1}^{m} L(\hat{y}^i, y^i) + \frac{\lambda}{2m} \sum_{j=1}^{n} W_j^2 \]

where \( \lambda \) is the regularization parameter.

Figure 2: Basic layout of machine learning algorithms

### 2.1.3 Other Algorithms in Machine Learning

The figure 2 shown above represents the traditional layout of the simple regression and logistic regression algorithms. While the regression and classification methods fall under the parametric algorithms of supervised learning, there are various other algorithms such as support vector machines, kernels, neural networks and other unsupervised learning algorithms such as clustering, dimensionality reduction, and recommender systems which are widely used now in industries and by researchers towards composite materials development.

### 2.1.4 Optimization Algorithms

The role of the optimization algorithm lies in obtaining the weight parameters \( W \) and \( b \), which will minimize the cost function \( J \). The Gradient Descent Algorithm is the most commonly used algorithm for the logistic regression. There are many optimization algorithms, such as L-BFGS, Adam algorithm which are used in neural network architectures. The optimization is the critical process in the machine learning algorithms, as the weights or parameters which are obtained as an output of these algorithms are used to make the predictions for the new data.

### 2.2 Neural Networks

Artificial neural networks (ANN) have been considered as the state of art technique for modeling and predicting non-linear system behavior. This non-linear behavior is obtained using hidden layers, which consist of units called neurons. The neurons in the hidden layers are connected to the input and output data by weights or parameters (\( W \) and \( b \)). The structure of an
artificial neural network is depicted as follows. Figure 3 shows the structure of N-Layer neural network with 3 inputs, predicting a single output.

![Image of N-Layer Neural Network]

Figure 3: Structure of Artificial Neural Network

### 2.2.1 Convolutional Neural Networks

Convolutional neural networks are a class of deep, feed-forward artificial neural networks that have been successfully used in analyzing visual imagery. CNNs use a variation of multilayer perceptron’s designed to require minimal preprocessing. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics. CNNs use relatively little pre-processing compared to other image classification algorithms. These networks learn the filters that in traditional algorithms were hand-engineered. CNNs are independent of the prior knowledge and they don’t need much human effort to process their features, as they train their filters automatically, which serves as an advantage of these networks. CNNs are typically used in composite materials engineering for image processing of fatigue damage images and other NDT testing image data.

### 2.2.2 Recurrent Neural Networks

Another major class of the artificial neural network is the recurrent neural network (RNN) where connections between nodes form a directed graph along a sequence. This connected tendency of RNN exhibits dynamic temporal behavior for a time sequence. Unlike traditional feed forward neural networks, RNNs use internal memory to process sequences of inputs. The RNNs have an additional stored state, which can be under direct control by the neural network. These stored states can also be replaced by a similar network, if that incorporates time delays or needs feedback loops. Such controlled states are referred to as gated state or gated memory, which belongs to long short-term memory’s (LSTM) and gated recurrent units (GRU).
3. DATA DRIVEN COMPOSITES

3.1 Design

Kasperkiewicz et al., concluded that the techniques of AI (Artificial Intelligence) can be applied successfully to analyze databases on the composition, on the properties, and on the applied experimental techniques, in testing of engineering materials. They developed the test data from experiments concerning properties of conventional concrete, of HPC (High Performance Concrete) and of SFRC (Steel Fibre Reinforced Concrete) and applied these AI solution methods: Fuzzy ARTMAP and aiNet (ANNs programs); aq18 (ML program); PCA (Principal Components Analysis); conventional Regression Analysis and observed that these procedures gave competitive results. [1]

Michopoulos et al., demonstrated the data driven application of design optimization methodologies for the determination of the constitutive response of composite materials with or without damage. They formulated the objective functions by expressing the difference between the experimentally observed behavior of composites and the simulated behavior of finite element analysis. [2]

Torbaghan, M., et al., developed artificial intelligence algorithms to design the composite materials based on the structural requirements, local structural stress and local material properties. They suggested that the mathematical models as well as finite element analysis and heuristic knowledge must be applied in order to find suitable materials and construct composite materials. [3]

Sticklen, J., et al., combined the experience of the seasoned designer’s protocols for composites and the fundamental studies of the materials involved to generate an artificial intelligence algorithm which will produce new designs. [4]

3.2 Analysis

Srivatsava et al., developed algorithmic frameworks for predicting properties of composites to aid in the meta-modelling process and in advancing the automated extraction of useful information from simulated data to make predictions at scales that are currently inaccessible. Their frameworks designed are applicable to virtually any new composite. They have been designed in a way that allows instantaneous recommendation of composite materials and prediction of their aggregate properties. [5]

3.3 Quality Inspection

Recent developments in the Machine learning and Neural Network algorithms, the usage of the neurons (brain-like structures) in supervised learning has developed a presence in the field of structural health monitoring and non-destructive testing.

Angelo, G D., and Rampone, S., proposed a model to classify the aerospace structure defects detected by eddy current non-destructive testing. They used decision trees, neural networks and the naïve Bayes models for the classification. [6]
Trtnik, G., et al., modelled an artificial neural network to predict concrete strength using the ultrasonic pulse velocity measurement. [7]

Worden, K., et al., discussed widely on the data driven approaches to monitor structures. They suggest that if the conditions are favorable, machine learning algorithms can be applied with great effect on damage identification. They insist, ‘favorable’ means that the data are available in order to train the machine learning diagnostics, and even for a model-driven approach, one will need appropriate data for model validation. [8]

3.4 Testing and Structural Health Monitoring

Loutas et al., developed a prognostic data driven method to predict the remaining useful life of composite materials under fatigue loading based on the acoustic emission results. They developed two machine learning algorithms – Non-Homogeneous Hidden Semi Markov Model and Bayesian feed forward neural networks, and further concluded that this NHHSN model outperformed the neural network model. [9]

Choi et al., developed a power law model and an ANN model to predict the split growth in notched AS4/3501-6 graphite/epoxy quasi-isotropic laminates under tension dominated fatigue. The ANN model worked better for them, which is depicted in figure 4. [10]

4. CHALLENGES AND PATH FORWARD

4.1 Challenges

4.1.1 Nature of the Input Data

For building a successful artificial intelligence models, the input data plays a significant role. Models developed with known physics based data, where the training data agrees with the
physics based relations always results in a successful model and the output of the model can be used for any particular application.

In certain cases, the physics based data is not available, and hence we have to depend on data-driven approaches. In such cases, these data driven approaches can also lead to the development of relations which can motivate and guide us to understand the physics. One such application is by using an ensemble learning algorithm, such as random forests algorithms, where we learn to know the importance of each input feature on the output is predicted. These important weights (depth of features) help us derive the relation of the model developed.

For “health monitoring,” as we discussed above, we need to recover and use data that contain information about the current integrity and prospective performance during the entire life of the material or component. This is a special challenge early in the life of the component when health is more likely reduced by events that are discrete in time and which create small but significant changes in the “health” of the material or structure. Specifically, how do we recover meaningful data early in the life of the component, not just near the end of life? For example finding and monitoring the length of a major crack or defect in the structural material (typically interpreted with fracture mechanics analysis) is the current certification data used by the FAA. But such cracks cannot be found until nearly the end of life, so instead of health monitoring we are only able to do “death monitoring.” Methods that provide meaningful material and system state data early in life are essential. Some show promise. Reifsnider, Raihan, et al. have developed material state assessment methods based on dielectric properties measured through the thickness. An example of such data is shown in Figure 5. [11-13]
That figure shows the stress-strain response of a woven cross ply laminate coupon tested at 45 degrees to the fiber axes, and also shows the through-thickness data for dielectric permittivity measured during the quasi-static loading.[11] The dielectric data show significant changes during the entire loading process, from the onset of minor cracking to the development of the final failure sequence and fracture. Those data can also be interpreted to show “the beginning of the end,” i.e., to identify the point in time when the rate of change in material state begins to increase, signifying the beginning of accelerated change in the material and system state leading up to the failure event that ends the life of the element.

A demonstration of that capability is shown in Figure 6, illustrating the point of this part of our discussion. By plotting the change in slope of the load-strain and the load-capacitance data curves in Figure 5, a clear minimum is identified. This critical point defines the “beginning of the end” of the system being monitored, an especially important and often essential element of the data-based method. Hence, more generally, the data selected for a data-centric AI method must include the information which, in fact, controls the essential features and behavior of the system being modeled, including the changes in those system states.

The recent death of a pedestrian struck by an autonomous vehicle driven by AI algorithms has highlighted the obvious need for uncertainty quantification in data driven procedures and methods [15]. Uncertainty quantification including the statistical significance of data points used for training and control, their associated confidence bounds and limits, and the propagation through the algorithm of those error features is critical features is essential to the qualification of data for AI analysis and system control.
4.1.2 Uncertainty of the prediction

In our previous research work, a Neural Network structure was defined and demonstrated that was successful in learning to predict notched strength, based on training with a data set generated from an average strength concept requiring three material constants. Excellent predictive fidelity was demonstrated from the models developed, as shown by the R² accuracy in figure 7.

![Figure 6: Comparison of the second variation (variation from linear slope) of the stored strain energy and dielectric capacitance.](image)

![Figure 7: R² Score of the Machine Learning Models developed from physics based equation [16]](image)

The statistical uncertainty of the artificial intelligence models being developed plays a major role when we need to predict the defect features, remaining life, strength of the composite material, or a new design for the composites structures. The statistical uncertainty is relatively negligible when the data driven model is coupled with a physics based model.
4.2 Path Forward

While Artificial Intelligence technologies are widely used in non-destructive testing and structural health monitoring of composites, there hasn’t been much research done on the Manufacturing part of composites with Artificial Intelligence. Hence there is a need for involvement of AI in the field of manufacturing, which could help us reduce the manufacturing cost and develop more new advanced manufacturing methods. These manufacturing methods should be based on the data of the composites.

As per the literature discussed above or the NDT, Tensile and Fatigue Testing, and other SHM techniques, the data collected are used in the data driven approaches, to build successful AI models, but the “missing physics” is not considered into effect.

The Institute of Predictive Performance Methodologies (IPPM) at the University of Texas at Arlington Research Institute have developed a method of using Broadband Dielectric Spectroscopy (BbDS) to detect the strength and integrity of composite structures, adhesive bonds in composite components. This is achieved by measuring dielectric properties of the composite structure or adhesively bonded joint, and showing direct relationships to the strength of the composite or adhesively bonded structure, and also showing unique links between changes in the bulk dielectric response and remaining strength and life.

Our recent research work at IPPM, established a nondestructive evaluation method by using machine learning techniques which could detect and predict the defects in the composite materials based on the initial dielectric properties obtained from BbDS.[17]

Figure 8: Artificial Intelligence combined with Physics Model for Real Time Decision Making

Based on the recent work, we postulate a hybrid approach of combining physics based models with the artificial intelligence models, which can relate in the real time learning and real time decision making, with very low uncertainty. A schematic diagram is shown in figure 8 which explains this hybrid approach concept. The recurrent neural networks, sequence models can be developed, which follows a time series or sequential data, implementing the stress-strain, fatigue life data, to predict life of the composite structures based on the initial response of the sequence models and initial dielectric properties.
5. CONCLUSIONS

A review of the artificial intelligence technologies used in the design and analysis of composites was analyzed. The challenges concerning the uncertainty of the models being developed, to predict the missing physics in the data are discussed. Also, a new hybrid methodology of combining artificial intelligence models with a physics based approach is proposed which uses the dielectric response data as the material state variables for the models.

6. REFERENCES


