Journal of Neuromorphic Intelligence

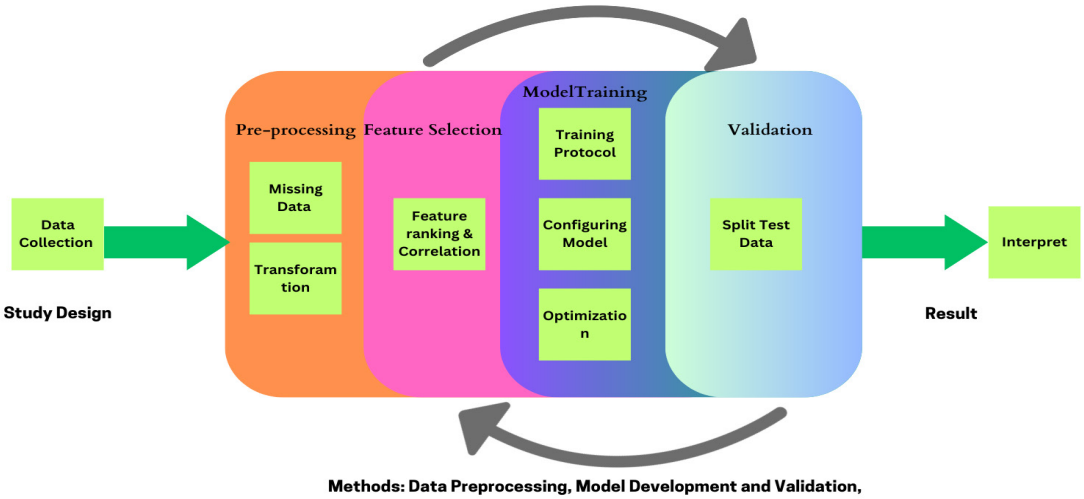
RESEARCH ARTICLE

VOLUME 01 | JULY 2024

Data-Efficient Machine Learning for in-situ Curing-aided Additive Manufacturing

Prashant Dhakal[1], Ruochen Liu[3], Jae Gwang Kim[2], Aolin Hou[2], Xiaofei Wu[1], Shiren Wang[1,2,*]

1 Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, USA
2 Department of Material Science and Engineering, Texas A&M University, College Station, TX, USA
3 Department of Material Science and Engineering, Beijing University of Technology, Bejing, China
** Corresponding author: s.wang@tamu.edu*



Received: 20 April 2024

Accepted: 6 June 2024

Published: 10 July 2024

Check for updates

ABSTRACT

In-situ curing assisted additive manufacturing (AM) of thermoset composites received numerous attentions and accurate prediction of the degree of cure (DOC) in thermosets is essential to online tailor process parameters for controlling the resultant quality and performance. Traditional methods for measuring DOC, such as Differential Scanning Calorimetry (DSC), Dynamic Mechanical Analysis (DMA), and Dielectric Analysis (DEA), are destructive or invasive, and difficult for a online test. In this paper, we present a data-driven approach to predicting DOC using machine learning. Our approach overcomes the limitations of traditional methods by considering localized variations, adapting to complex curing kinetics, and directly predicting DOC. We compare two machine learning algorithms: support vector classifier (SVC) and random forest, using a small dataset. SVC achieved an accuracy of 81%, a precision of 0.86, and a recall of 0.81, while random forest achieved an accuracy of 78%, a precision of 0.83, and a recall of 0.78. This demonstrates the feasibility of data-driven DOC prediction in AM, paving the way for enhanced manufacturing processes.

Keywords: ***Additive manufacturing, in situ curing, degree of cure, machine learning, quality control, material design***

1. INTRODUCTION

In the in-situ curing assisted additive manufacturing (AM) processes, such as direct ink writing (DIW), accurate degree of cure (DOC) prediction is vital for optimizing the printing process, ensuring print- ability, achieving desired mechanical properties, maintaining dimensional stability, and to obtain overall quality products. This is because the DOC can vary throughout the build volume, depending on factors such as the print speed, temperature, and layer thickness. It also guides post-processing steps such as post-curing. The level of curing indicates the extent of polymerization in the polymer material. Generally, a higher level of curing is beneficial for enhancing the mechanical characteristics of the composite.[1] However, accurate prediction faces challenges due to the nonlinearity of chemical reactions, limited access to the curing zone for real-time monitoring, process variability, and the interplay between material properties, ink deposition, and curing conditions [2]. The curing process is influenced by several factors, including the material properties, the print parameters, and the ambient environment. As a result, it is difficult to develop a single model that can accurately predict the DOC for all AM processes and materials. Despite the challenges, accurate DOC prediction is important for several reasons. First, it can help to ensure that the final product meets the desired properties. Second, it can help to avoid defects, such as voids and cracks [1]. Third, it can help to optimize the print parameters, such as the print speed and temperature.

Various cure monitoring techniques are available for composites, with Differential Scanning Calorimetry (DSC) and Dynamic Mechanical Analysis (DMA) being the most used methods. In a dynamic heating DSC experiment [3], it is possible to determine the instantaneous glass transition temperature (Tgmom) and the residual heat generated during the cross-linking reaction. Dynamic heating refers to a non-isothermal ramp temperature program. The degree of cure can be assessed by correlating the residual heat in a partially cured sample with the total heat of cure in a fully cured sample. However, the DSC method has a disadvantage when dealing with partially cured materials, as the endothermic step at the glass transition often overlaps with the residual exothermic peak, making separation challenging [4]. Additionally, at higher degrees of cure, the residual heat is low, resulting in imprecise calculations of the degree of cure. DSC is also not suitable when dealing with large number of samples and volatile samples [5].

Dynamic Mechanical Analysis (DMA) is an often-employed technique for assessing curing progression and cure condition [6], [7]. The cross-linking reaction causes distinctive changes in mechanical properties, which DMA uses to provide useful information on the status of cure. It is capable of reliably analyzing materials with a high degree of cure and can quickly detect the glass transition. DMA is known for its sensitivity in measuring subtle transitions in polymers and is often the preferred method for determining the glass transition temperature (Tg) in highly cross-linked thermosetting resins, as other methods like DSC and Thermomechanical Analysis (TMA) may lack sufficient sensitivity [8]. While DSC and DMA offer valuable insights into parameters such as Tg, onset of cure, heat of cure, maximum rate of cure, completion of cure, and degree of cure, they do have limitations. These methods are destructive and can only be applied on a laboratory scale [9].

To address these limitations and minimize disruption to the manufacturing process, non- invasive and non-destructive methods are sought after. It is crucial that such methods provide real-time information about the cure reaction. One such method is Dielectric Analysis (DEA), which is widely used. The most common dielectric measurement system utilizes a parallel plate capacitor or an interdigital capacitor (IDC). However, a drawback of this method is that the dielectric sensors need to be embedded within the fabrics, which has been reported to adversely affect the mechanical performance of the parts [5].

Having completed a comprehensive literature review on the current approaches utilized for degree of cure prediction, it becomes apparent that the field is ripe for a paradigm shift towards the utilization of machine learning techniques. While the reviewed methods made significant contributions to understanding and determining the degree of cure, they are often limited by certain constraints. Traditional approaches rely on explicit mathematical models that require extensive empirical data and assumptions about the underlying curing mechanisms [10]. Additionally, these methods may struggle to capture complex nonlinear relationships and intricate patterns present in the curing process. Moreover, these methods are usually limited to in-lab setups hence are not feasible at an industrial production and processing level [9]. And finally, they greatly slow manufacturing times, and in some cases, damage the part [9]. In contrast, machine learning offers a data-driven alternative that can automatically extract valuable insights from large datasets without explicitly defined equations or assumptions. By transitioning to machine learning, researchers can leverage its ability to uncover hidden patterns, handle high-dimensional data, and adapt to nonlinearities, ultimately leading to more accurate and robust degree of cure predictions [11].

In this study, a predictive model was developed to predict the degree of cure. Firstly, different algorithms were compared and analyzed to check which performed the best. The impact of different factors on the accuracy of the model, such as the type of data used, and the complexity of the model, were also analyzed. The rest of the paper is organized into the following parts: Section 2 introduces the background knowledge of data driven machine learning models. Section 3 discusses the experimental setup and data collection. Section 4 contains the analysis and interpretation of the results. The paper closes with some conclusions and future directions in the last section

2. DATA DRIVEN MODELS

2.1 RANDOM FOREST

Random Forest (RF) is a powerful ensemble learning algorithm utilized in this study. It involves constructing numerous independent regression trees [12-14]. To obtain the ideal split at each node, the following objective function must be solved:

$$\min_{j,s}\left[\min_{c1}\sum_{x_i\in R_1(j,s)}(y_1-c1)^2+\min_{c2}\sum_{x_i\in R_2(j,s)}(y_1-c2)^2\right] \quad (1)$$

Let y_i represent the response variable of the ith sample x_j , where j ranges from 1 to p (p being the number of splitting variables). The cutting point is denoted as S. When the best split is performed, it results in two regions R_1 (j,s)={X | $X_j \le s$} and R_2 (j,s)= {X | $X_j \ge s$} .Here, X_j represents the jth splitting variable. The average of y_i values within region R_1 (j,s).c2=ave(y_i | x_i ϵ R_1 (j,s)) is denoted as c1=ave(y_i | x_i ϵ R_1 (j,s)), while R_1 (j,s).c2=ave(y_i | x_i ϵ R_1 (j,s))represents the average of y_i values within region R_2(j,s).
The process continues until the defined terminal condition has been met.

2.2 SVC

In support vector following optimization problem is solved [16,17]

Minimize

$$\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{\hat{}}(\xi_i+\xi_i^*) \quad (2)$$

Subject to

$$\begin{cases} y_i-\langle w,x_i\rangle-b\le\varepsilon+\xi_i \\ \langle w,x_i\rangle+b-y_i\le\varepsilon+\xi_i^* \\ \xi_i,\xi_i^*\ge 0 \end{cases}$$

In this context, the parameter C (> 0) determines the balance between the smoothness of the function, f(x)= w, ├ x +b┤ and the extent to which deviations beyond ε are permitted. To make the data separable, kernel functions are employed to transform the data into higher-dimensional space. [17,18].

2.3 XGB

XGBoost (Extreme Gradient Boosting) is a powerful engineering algorithm inspired by the Gradient Boosting Decision Tree (GBDT) algorithm, which was originally developed by Dr. Chen Tianqi from the University of Washington [19]. The core idea behind the GBDT algorithm is to guide the learning process of each tree in the direction of gradient descent. The learning process is given below:
For the model Fm(x) under the current iteration number of round m, it can be expressed as:

$$F_m(x)=F_{m-1}(x)+argmin\sum_{i-1}^{n}loss(y_i,F_{m-1}(x_i)+f(x_i))F_{m-1}(x) \quad (3)$$

F_(m-1) (x)is the model obtained by m-1 iteration, f(x) is the minimized loss function and xi,yi are training samples.
The algorithm further improves the loss function by gradient descent method and the vector moves towards the negative gradient direction of the loss function, it is shown below:

The new training set is obtained by the following equation and the new tree is obtained which is trained again.

$$v=-\left(\frac{\delta loss(y_1,F_{m-1}(x_1))}{\delta F_{m-1}(x_1)},\frac{\delta loss(y_2,F_{m-1}(x_2))}{\delta F_{m-1}(x_2)},\dots\dots,\frac{\delta loss(y_n,F_{m-1}(x_n))}{\delta F_{m-1}(x_n)}\right) \quad (4)$$

The XGBoost algorithm incorporates several engineering optimizations based on the traditional GBDT algorithm. These optimizations include Integration of a regularization term, Second-order Taylor expansion, and Sample and feature subsampling. These engineering optimizations implemented in XGBoost aim to enhance the algorithm's performance, robustness, and generalization ability, making it a highly effective tool for various applications.

$$\left\{x_i,\frac{\delta loss(y_i,F_{m-1}(x_i))}{\delta F_{m-1}(x_i)}\right\}_{i=1}^{n} \quad (5)$$

3. EXPERIMENTAL SETUP AND DATA COLLECTION

3.1 3D PRINTING

The micro-powder reactants were prepared by mixing powders of Ti, Si, and graphite in a molar ratio of 3:1:2. Specifically, the mixing of powders was performed in ethanol suspension in an ultrasonication bath for 1 h followed by mechanically stirring for 1 h. To remove the ethanol, the suspension was then dried at 85 °C under mild mechanical stirring overnight until ethanol content was not detectable with differential scanning calorimetry. The feedstock was then extruded using a 3D Potter model Micro 10 printer equipped with a 1mm nozzle, controlled digitally (Fig. 2a) . For the UV in-situ curing 3D printing process, an Omnicure S2000 device from Excelitas technologies, featuring two-way light guides, was utilized. To ensure accurate power delivery, the radiometer recorded the various power levels directed towards the sample, which were subsequently calibrated based on the distance between the output portal and the target. The various printing parameters used are given below as Table 1.:

Table 1: Process Parameters used for DIW printing

Process Parameters	
UV Power (W/cm2)	19, 18,17,16,15,14,13,12
Scan Speed(mm/ min)	20,40,60,80,100,120,140,160,180

3.2 DATA COLLECTION

Thermal analysis was conducted using a Differential Scanning Calorimeter (DSC) (Q20, TA Instruments) equipped with a CFL-50 cooling system. The DSC analysis involved the use of samples weighing between 50-60mg, which were placed in aluminum hermetic DSC pans and sealed. A programmed heating curve was applied, starting from 40 °C and reaching 350 °C at a heating rate of 5 °C/min for dynamic DSC scanning. The reaction enthalpy was determined by calculating the total area of heat flow after baseline correction. Additionally, the degree of cure and the rate of the degree of cure were evaluated in an isothermal scanning experiment , specifically at temperatures ranging from 125 to 145 °C. In Fig 3a, the first peak at 168.97°C represents an initial exothermic reaction or curing stage, while the second peak at 265.51°C indicates a secondary reaction or further curing process at a higher temperature. Together, these peaks highlight distinct stages of the exothermic curing or reaction process. To analyze temperature variations, an infrared (IR) camera (FLIR A325sc) was employed to capture the temperature profile during the printing process. The recorded IR images were then processed to extract the maximum temperature within the region of interest (ROI) for the printed sample.

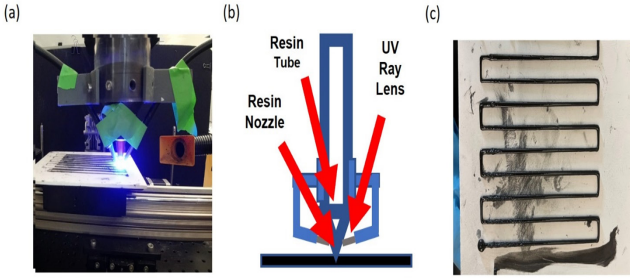


Figure 2: Photothermal induced in-situ curing. (a) Picture showing UV Ray curing process of the resin. (b) Detailed schematic of the 3D printer. (c) 3D printed sample

3.3 DATA PREPARATION AND MODEL TRAINING STRATEGY

In this study, a dataset consisting of 63 data points was utilized. The dataset comprised three features, namely Power, Speed, and Temperature, and one target variable, which represented the degree of cure. To assess the performance of the machine learning models, the dataset was split into training and testing subsets, with a ratio of (20-33)% allocated for testing purposes. To ensure reliable model evaluation, the k-fold cross-validation technique was employed for validation.
During the training phase, various machine learning algorithms were utilized, and grid search cross-validation was conducted to explore different combinations of training parameters. This process enabled the evaluation of multiple models and facilitated the selection of the best-performing model for each algorithm. The performance of the selected models was then compared. To assess the models' performance, the testing dataset, which was not involved in the training process, was employed. For classification problem, the degree of cure values was categorized into three distinct categories: "low cured" (values less than 0.4), "moderate cured" (values ranging from 0.4 to 0.6), and "high cured" (values greater than 0.6).

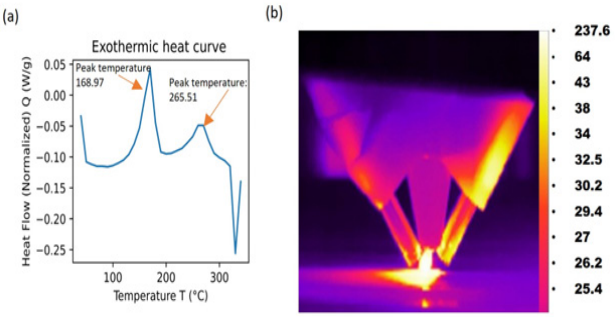


Figure 3: (a). DSC measurements for Sample 1 : exothermic heat curves at a rate of 5 °C/min,(b)infrared (IR) image analysis of sample while it's being cured

3.4 EVALUATION METRICS

In this study both regression and classification methods were examined. For regression, several evaluation metrics were employed

3.4.1 MEAN SQUARED ERROR (MSE)

MSE is a commonly used metric to measure the average squared difference between the predicted values and the actual values. It provides a quantitative measure of the overall accuracy of a regression model [23-25].

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2 \quad (6)$$

Xi is the predicted ith value and the Yi element is the actual ith value.

3.4.1 COEFFICIENT OF DETERMINATION(R^2)

R-squared is a statistical measure that represents the proportion of the variance in the dependent variable (target variable) that is predictable from the independent variables (features) in a regression model [26-28].

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(X_i - Y_i)^2}{\sum_{i=1}^{m}(\bar{Y} - Y_i)^2} \quad (7)$$

Xi is the predicted ith value and the Yi element is the actual ith value. And is the mean of true values.

3.4.1 MEAN ABSOLUTE PERCENTAGE ERROR (MAPE)

It calculates the average percentage difference between the predicted values and the actual values, relative to the actual values. MAPE provides an understanding of the magnitude of the prediction errors in relation to the actual values [29-30].

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{Y_i - X_i}{Y_i}\right| \quad (8)$$

3.4.4 CLASSIFICATION METRICS

Precision, recall, and f-score are commonly employed as evaluation metrics in multi-class classification scenarios [31-33]. In the context of a multi-class problem with class Ci, where i ranges from 1 to l, these metrics can be computed using values derived from the confusion matrix, including true positives (tp), true negatives (tn), false negatives (fn), and false positives (fp). Accuracy provides an estimate of the proportion of correctly predicted instances overall, offering a general assessment of classification performance. Precision focuses on the true positive predictions. And evaluates the model's ability to accurately identify positives. Conversely, recall measures the false negatives and assesses the model's capability to correctly classify instances belonging to a certain class. The F1 score is the harmonic mean of precision and recall, striking a balance between the two metrics and disregarding the true negative (tn) count. In this study, the "macro-averaging" strategy is utilized, which assigns equal weight to each class, ensuring an unbiased evaluation across all classes [34].

$$Precision = \frac{\sum_{i=1}^{l} \frac{t_{pi}}{t_{pi} + f_{pi}}}{l} \quad (9)$$

$$Accuracy = \frac{\sum_{i=1}^{l} \frac{t_{pi} + t_{ni}}{t_{pi} + f_{ni} + f_{pi} + t_{ni}}}{l} \quad (10)$$

$$Recall = \frac{\sum_{i=1}^{l} \frac{t_{pi}}{t_{pi} + f_{ni}}}{l} \quad (11)$$

$$F_1 = \frac{2Precision.Recall}{Precision + Recall} \quad (12)$$

4. RESULTS AND DISCUSSION

The study involved gathering data from the experiment, resulting in a dataset consisting of 63 samples. Each data set included the experimental conditions (features). Since machine learning models necessitate a significant volume of data, caution was exercised regarding the feasibility of employing advanced techniques such as neural networks [35-38]. As a result, two algorithms, namely RF, and SVC [39-40] were chosen, which are suitable for smaller datasets. With the goal of predicting the degree of cure, three regression models were built. The results obtained during the training and testing stages of the proposed method are shown below. The examination of the three key visualizations, namely Spearman's coefficient, feature importance, and class distribution, was initiated. Figure 4 presented these visualizations, which provided valuable insights into the relationships between variables, the significance of features, and the distribution of classes within the dataset. In particular, the first image, 4(a), displayed Spearman's coefficient, which depicted the correlation between different variables in the dataset. Spearman's coefficient was utilized to measure the strength and direction of monotonic relationships between variables, without assuming linearity [41-42][44]. The second image, 4(b), showcased feature importance, offering a comprehensive understanding of the contribution made by each feature towards the prediction task [43]. The computation of the feature importance score involved the utilization of tree-based models, such as Random Forest. Lastly, the third image, 4(c), portrayed class distribution, illustrating the distribution of classes or categories within the dataset. The visualization of class distribution enabled the gaining of insights into the balance or imbalance of classes, which could influence model training and evaluation.

The correlation values provide insights into the relationships between the features (temperature, speed, and power) and the target variable (cure). The temperature feature demonstrates a positive correlation coefficient of 0.45, indicating a moderately strong positive relationship with the cure. This suggests that as the temperature increases, there is a tendency for the cure to also increase. On the other hand, the speed feature exhibits a negative correlation coefficient of -0.44, suggesting a moderately strong negative relationship with the cure. This implies that as the speed increases, the cure tends to decrease. Interestingly, the power feature shows a weak positive correlation coefficient of 0.079, indicating a minimal relationship with the cure. Moving on to the feature importance scores, which provide insights into the relative importance of each feature in predicting the cure. The temperature feature emerges as the most influential feature with a score of 0.53, indicating its strong impact on the prediction of

Table 2:Statistical analysis of all the parameters involved in machine learning model of degree of cure prediction

Statistics	Parameters			
	Power	Speed	Temperature	Cure
Count	63	63	63	63
Mean	81.76	95.23	219.86	0.418
Standard Deviation	9.93	48.68	90.47	0.161
Min	67	20	62.2	0.089
25%	75	60	135.59	0.313
50%	79	100	218.2	0.431
75%	90.5	140	297.15	0.534
Max	98	200	360.14	0.989

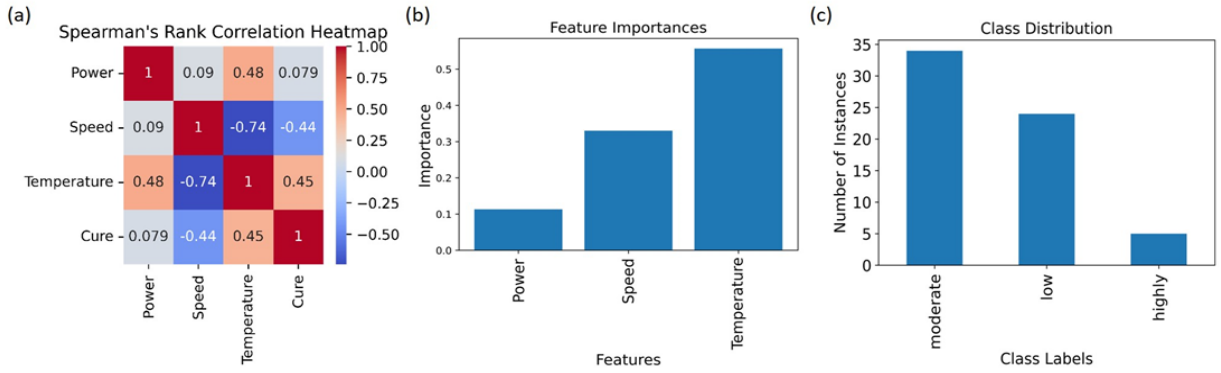


Figure 4 (a) Spearman's rank correlation relationship between input and output parameters, (b) Feature ranking, (c) Class distribution.

the cure. The speed feature follows with a score of 0.35, indicating a considerable contribution to the prediction task. Lastly, the power feature demonstrates the lowest importance score of 0.12, implying a relatively minor influence on the prediction of the cure. Figure 4 (c) shows that the dataset is imbalanced, with moderately cured having higher instances of 34, low cured class with 24 instances and highly cured with 5 instances. Presence of such imbalanced dataset requires modifying the algorithm and choosing the appropriate evaluation metrics.

Random Forest Regressor and XGB reg was built and evaluated. The parameter for the model are given below:

Table 3: Hyperparameters for Regression models

Parameters	XGB REg	RF Reg
n_estimators	9	80
max_sample	-	-
max_depth	5	5
max_feature	-	1
colsmaple_ bylevel	1	-
colsmaple_bytree	0.7597	-
learning_rate	0.3513	-
Base_estimator	-	-
min_samples_ split	-	2
min_child_weight	0.36957	-
subsample	0.9586	-
reg_alpha	0.335	-
reg_lambda	0.1123	-
min_samples_leaf	-	3
random_state	5	5

R2, MSE and MAPE evaluation metrics were used to assess the performance of the model on the test data (20-33% of the dataset). The scatter plot and the obtained values are given in Table 4,

Table 4: Model's evaluation on Test data

Model	MSE	R2	MAPE
XGB Reg	0.00384	0.56	4.3%
RF Reg	0.004	0.53	5%

Starting with the coefficient of determination (R^2), which provides an indication of how well the models capture the variance in the test data, we obtained values of 0.56 and 0.53 for the XGB Regressor, and Random Forest Regressor, respectively. These values suggest that the models explain approximately 56% to 53% of the variance in the test data. Moving on to the mean squared error (MSE), which measures the average squared difference between the predicted and actual values, we obtained values of 0.00384 and 0.004 for the XGB Regressor, and Random Forest Regressor, respectively. The lower the MSE, the better the model's predictive accuracy. In this case, all three models demonstrate relatively low MSE values, suggesting that they can make accurate predictions with minimal error. Additionally, we evaluated the mean absolute percentage error (MAPE) to assess the models' performance in terms of relative errors. The XGB Regressor achieved a MAPE of 4.3%, and Random Forest Regressor achieved MAPE values of 5%, respectively. The MAPE indicates the average percentage deviation of the predictions from the actual values. Lower MAPE values indicate a higher level of accuracy. In this case, the XGB Regressor exhibits the lowest MAPE, followed by and Random Forest Regressor. This suggests that the XGB Regressor model outperforms the other two models in terms of minimizing relative prediction errors. The XGB Regressor is a gradient boosting algorithm that combines the strengths of decision trees and gradient descent optimization. It is known for its ability to handle small datasets effectively. With its ensemble of weak learners, it can learn complex relationships and adapt to the data's non-linear nature. Given the limited dataset size, the XGB Regressor's capability to

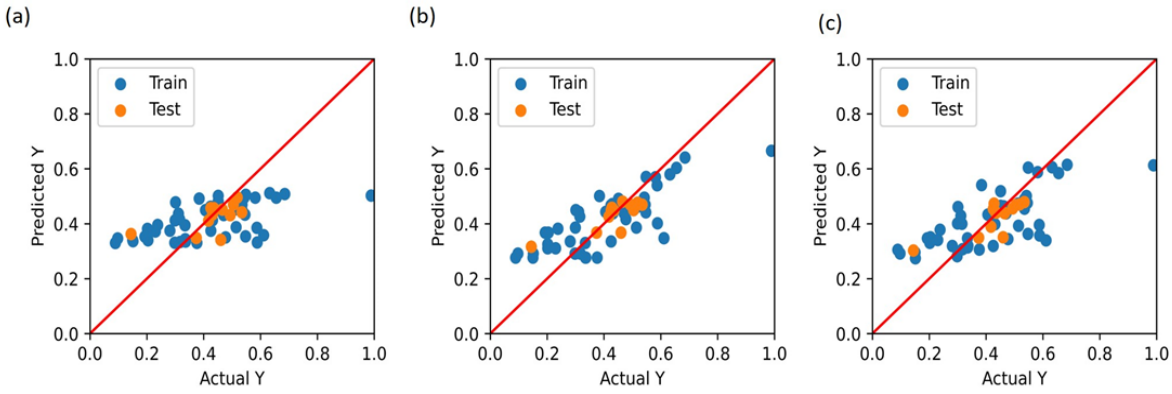


Figure 5 : Scatter Plots of Training data and Testing data for degree of cure prediction yield by (a) Bagging Regressor, (b) XGB Regressor, (c)) RF Regressor

In the context of additive manufacturing (AM), the accurate prediction of the degree of cure held utmost importance in this study. However, it became evident from the results that the regression models fell short of expectations and failed to accurately predict the degree of cure values. The authors hypothesized that the reduced accuracy of the regression models could be attributed to the limited size of their dataset. The limited size of dataset posed a hindrance to the effective training of the models.

To address this issue, an alternative approach was devised by transforming the task of predicting the degree of cure from a regression problem into a classification problem. This involved the categorization of the degree of cure values into three distinct categories: “low cured” (values less than 0.4), “moderate cured” (values ranging from 0.4 to 0.6), and “high cured” (values greater than 0.6). By adopting this classification framework, we were able to utilize three machine learning algorithms to train models specifically designed for predicting the degree of cure based on the experimental conditions. Classification model SVC was built. The parameters of the model are givenn Table 5.

The testing results of the three models on the test data (20-33% of the dataset) are presented in the confusion matrices as shown in figure 6. Figure 6(a) SVC classifier correctly classifies all instances from class “low”, misclassifies 4 from class “moderate” and correctly classifies instances from class “high”. Intuitive analysis of the reason could be the imbalance in dataset resulting in the misclassification.

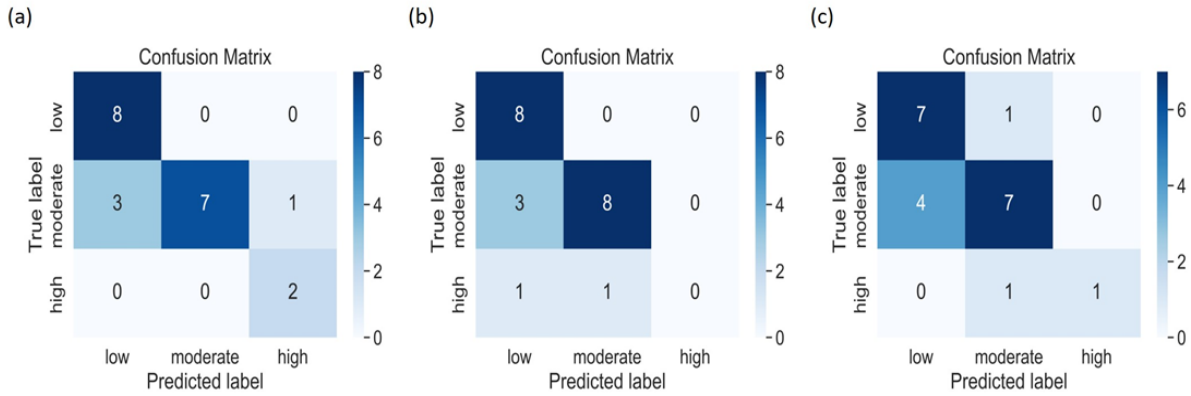


Figure 6: Confusion matrix (a) SVC, (b) Ada boost classifier, (c) XGB classifier

Due to the presence of an imbalanced dataset, evaluating the model's performance solely through a confusion matrix proved to be challenging. Therefore, an exploration of alternative metrics was conducted, and the corresponding results are presented in the table 6.

Table 5: Hyperparameters for classification models

Parameter	SVC
n_estimators	-
max_depth	-
max_leaves	-
max_feature	-
learning_rate	-
Base_estimator	-
objective	-
Gamma	scale
class_weight	balanced
probability	true
Random_state	42

The performance of SVC was evaluated based on precision, recall, specificity, accuracy, and F-1 score. Overall, the SVC model outperformed the other models in most metrics, demonstrating higher precision (0.86) and recall (0.81). This suggests that the SVC model had a better ability to correctly identify instances belonging to specific classes. It also achieved a higher specificity (0.91), indicating a strong capability to accurately classify negative instances. The SVC model's accuracy (81%) was also the highest among the three models, further solidifying its performance.

Accurate prediction of the degree of cure in thermosets is a crucial task in additive manufacturing (AM). Traditional approaches, such as Differential Scanning Calorimetry (DSC), Dynamic Mechanical Analysis (DMA), and Dielectric Analysis (DEA), have been widely employed for this purpose. However, there is a need for an efficient, non-destructive ad non-invasive approach in measuring the Degree of cure. For instance, DSC measurement on a large amount of sample is very time consuming and limited to only laboratory scale. They also rely on predefined mathematical models or assumptions about the curing process which sometimes is not the case for some systems.

Table 6: Hyperparameters for classification models

Model	Precision	Recall	Specificity	Accuracy	F-1
SVC	0.86	0.81	0.91	81	0.80

In this manuscript, a data-driven approach for Degree of cure measurement was presented. Both regression and classification models were examined. Notably, the classification models, namely SVC exhibited significant advantages over the regression models. While the regression models failed to accurately predict the degree of cure values, the classification models successfully categorized the degree of cure into three distinct categories: "low cured," "moderate cured," and "high cured." Furthermore, the evaluation metrics of the classification models emphasized their performance in predicting the degree of cure categories. For instance, the SVC model achieved high precision, recall, specificity, accuracy, and F-1 score, indicating its effectiveness in accurately classifying thermosets into the appropriate degree of cure categories. However, there are several factors that affect this study. First, limitation of dataset. The data set used in this study was relatively small. This could have led to overfitting of the models. Another limitation is the reliance on a specific machine learning algorithm, such as XGB Classifier, SVC, and AdaBoost, for classification modeling. While these algorithms have shown promising results in this study dealing with small dataset, there may be alternative algorithms or ensemble techniques that could yield even better performance. And finally, the third factor is the quality of the training process. This includes answering the following questions: What is the best strategy for determining the number and importance of features?What is the correct number of trees to generate? How should error rates be evaluated?

5. CONCLUSION

In this paper, a data-driven approach was developed using machine learning algorithms to accurately predict the degree of cure (DOC) in the in-situ curing aimed AM process of thermosets. The performance of five different models were compared and it was found that the support vector classifier (SVC) model achieved the highest accuracy, precision, and recall. This data-driven approach overcomes the limitations of traditional methods by considering localized variations, adapting to complex curing kinetics, and directly predicting the degree of cure. The results demonstrate the superiority of our approach in accurately predicting DOC and provide valuable insights for quality control and process optimization in AM.

Despite the success of this data-driven approach, there are still avenues for further research and improvement. First, expanding the dataset size could enhance the accuracy and generalization capabilities of the models. A larger dataset would provide more diverse samples and improve the models' ability to capture the complex relationships within the curing process. Additionally, exploring other machine learning algorithms or ensemble techniques may yield even better performance and robustness. Moreover, investigating the impact of different features or feature engineering techniques could provide additional insights into the curing process and potentially enhance the models' predictive capabilities.

Furthermore, it would be valuable to evaluate the performance of our approach on different thermoset materials and manufacturing conditions. This would help assess the generalizability and applicability of the models across various scenarios. Moreover, incorporating real-time monitoring and feedback mechanisms into the predictive models could enable adaptive control strategies and real-time adjustments during the AM process.

In conclusion, our research demonstrates the effectiveness of a data-driven approach using machine learning for predicting the degree of cure in thermosets for AM applications. This approach has

the potential to revolutionize quality control and process optimization in AM, leading to improved manufacturing outcomes and enhanced material design. Future work should focus on expanding the dataset, exploring alternative algorithms, and evaluating the approach on different materials and manufacturing conditions to further advance the field of predictive models for thermoset curing in AM.

REFERENCES

1. Ferracane, J. L., Mitchem, J. C., Condon, J. R., & Todd, R. (1997). Wear and marginal breakdown of composites with various degrees of cure. Journal of dental research, 76(8), 1508-1516.
2. Yang, Y., Li, L., & Zhao, J. (2019). Mechanical property modeling of photosensitive liquid resin in stereolithography additive manufacturing: Bridging degree of cure with tensile strength and hardness. Materials & Design, 162, 418-428.
3.] Höhne, G. W. H., Hemminger, W., & Flammersheim, H. J. (2003). Differential scanning calorimetry (Vol. 2, pp. 9-30). Berlin: Springer
4. Stark, W., Jaunich, M., & McHugh, J. (2013). Cure state detection for pre-cured carbon-fibre epoxy prepreg (CFC) using Temperature-Modulated Differential Scanning Calorimetry (TMDSC). Polymer testing, 32(7), 1261-1272.
5. Hardis, R., Jessop, J. L., Peters, F. E., & Kessler, M. R. (2013). Cure kinetics characterization and monitoring of an epoxy resin using DSC, Raman spectroscopy, and DEA. Composites Part A: Applied Science and Manufacturing, 49, 100-108
6.] Ehrenstein, G. W., Riedel, G., & Trawiel, P. (2012). Thermal analysis of plastics: theory and practice. Carl Hanser Verlag GmbH Co KG
7. Menard, K. P., & Menard, N. (2020). Dynamic mechanical analysis. CRC press.
8.] Pascault, J. P., & Williams, R. J. (2013). Thermosetting polymers. Handbook of Polymer Synthesis, Characterization, and Processing, 519-533
9.] García-Manrique, J. A., Marí, B., Ribes-Greus, A., Monreal, L., Teruel, R., Gascón, L., ... & Marí-Guaita, J. (2019). Study of the degree of cure through thermal analysis and Raman spectroscopy in composite-forming processes. Materials, 12(23), 3991.
10. Bernath, A., Kärger, L., & Henning, F. (2016). Accurate cure modeling for isothermal processing of fast curing epoxy resins. Polymers, 8(11), 390.
11. Gao, C., Qiu, J., & Wang, S. (2022). In situ curing of 3D printed freestanding thermosets. Journal of Advanced Manufacturing and Processing, 4(3), e10114.
12. L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
13.] A. Liaw, M. Wiener, Classification and regression by randomForest, R News 2 (3) (2002) 18–22.
14. Li, Z., Zhang, Z., Shi, J., & Wu, D. (2019). Prediction of surface roughness in extrusion-based additive manufacturing with machine learning. Robotics and Computer-Integrated Manufacturing, 57, 488-495.
15. Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 55 (1) (1997) 119–139
16.] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
17. A. Smola, V. Vapnik, “Support vector regression machines, Adv. Neural Inf. Process. Syst. 9 (1997) 155–161.
18. A.E. Hoerl, R.W. Kennard, Ridge regression: biased estimation for nonorthogonal problems, Technometrics 12 (1) (1970) 55–67.
19. Deng, J., Xu, Y., Zuo, Z., Hou, Z., & Chen, S. (2019). Bead geometry prediction for multi-layer and multi-bead wire and arc additive manufacturing based on XGBoost. In Transactions on Intelligent Welding Manufacturing: Volume II No. 4 2018 (pp. 125-135). Springer Singapore.
20. Buhlmann P, Yu B (2002) Analyzing bagging. Ann Stat 30:927–61
21. Breiman L (1996a) Bagging predictors. Mach Learn 24:123–40
22. Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems, 9, 181-199.
23.] Sammut, C., & Webb, G. I. (2010). Mean squared error. Encyclopedia of Machine Learning, 653.
24. David, I. P., & Sukhatme, B. V. (1974). On the bias and mean square error of the ratio estimator. Journal of the American Statistical Association, 69(346), 464-466.
25. Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, e623.
26. Di Bucchianico, A. (2008). Coefficient of determination (R 2). Encyclopedia of statistics in quality and reliability.
27.] Barrett, G. B. (2000). The Coefficient of Determination: Understanding r squared and R squared. The Mathematics Teacher, 93(3), 230-234.
28. Quinino, R. C., Reis, E. A., & Bessegato, L. F. (2013). Using the coefficient of determination R2 to test the significance of multiple linear regression. Teaching Statistics, 35(2), 84-88.
29. De Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F. (2016). Mean absolute percentage error for regression models. Neurocomputing, 192, 38-48.
30. Ren, L., & Glasure, Y. (2009). Applicability of the revised mean absolute percentage errors (MAPE) approach to some popular normal and non-normal independent time series. International Advances in Economic Research, 15, 409-420.
31.] Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. International journal of data mining & knowledge management process, 5(2), 1.
32. Foody, G. M., & Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. IEEE Transactions on geoscience and remote sensing, 42(6), 1335-1343.
33. Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: an overview. arXiv preprint arXiv:2008.05756.
34. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information processing & management, 45(4), 427-437.
35. Lee, J. W.; Park, W. B.; Lee, J. H.; Singh, S. P.; Sohn, K. S. A deep-learning technique for phase identification in multiphase inorganic compounds using synthetic XRD

powder patterns. Nat Commun 2020, 11 (1), 86. DOI: 10.1038/s41467-019-13749-3.
36. Azimi, S. M.; Britz, D.; Engstler, M.; Fritz, M.; Mucklich, F. Advanced Steel Microstructural Classification by Deep Learning Methods. Sci Rep 2018, 8 (1), 2128. DOI: 10.1038/s41598-018-20037-5.
37. Ju, L.; Lyu, A.; Hao, H.; Shen, W.; Cui, H. Deep Learning-Assisted Three-Dimensional Fluorescence Difference Spectroscopy for Identification and Semiquantification of Illicit Drugs in Biofluids. Analytical Chemistry 2019, 91 (15), 9343-9347. DOI: 10.1021/acs.analchem.9b01315.
38. Ziatdinov, M.; Dyck, O.; Maksov, A.; Li, X.; Sang, X.; Xiao, K.; Unocic, R. R.; Vasudevan, R.; Jesse, S.; Kalinin, S. V. Deep Learning of Atomically Resolved Scanning Transmission Electron Microscopy Images: Chemical Identification and Tracking Local Transformations. ACS Nano 2017, 11 (12), 12742-12752. DOI: 10.1021/acsnano.7b07504.
39.] Liaw, A.; Wiener, M. Classification and regression by randomForest. R news 2002, 2 (3), 18-22.
40.] Cortes, C.; Vapnik, V. Support-vector networks. Machine Learning 1995, 20 (3), 273-297. DOI: 10.1007/BF00994018.
41. Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaestiones geographicae, 30(2), 87-93.
42. Myers, L., & Sirois, M. J. (2004). Spearman correlation coefficients, differences between. Encyclopedia of statistical sciences, 12.
43. Jong, K., Mary, J., Cornuéjols, A., Marchiori, E., & Sebag, M. (2004). Ensemble feature ranking. In Knowledge Discovery in Databases: PKDD 2004: 8th European Conference on Principles and Practice of Knowledge Discovery in Databases, Pisa, Italy, September 20-24, 2004. Proceedings 8 (pp. 267-278). Springer Berlin Heidelberg.
44. Liu, R., Kim, J. G., Dhakal, P., Li, W., Ma, J., Hou, A., ... & Wang, S. (2023). Neuromorphic properties of flexible carbon nanotube/polydimethylsiloxane nanocomposites. Advanced Composites and Hybrid Materials, 6(1), 14.

COMPETING INTERESTS

The authors declare no competing interests.

ACKNOWLEDGMENTS

The authors are grateful for the funding support from Texas A&M University