A machine learning approach to characterise fabrication porosity effects on the mechanical properties of additively manufactured thermoplastic composites

Udu, A. G., Osa-Uwagboe, N., Adeniran, O., Aremu, D., Khaksar, M. G. & Dong, H.

Published PDF deposited in Coventry University's Repository

Original citation:

Udu, AG, Osa-Uwagboe, N, Adeniran, O, Aremu, D, Khaksar, MG & Dong, H 2024, 'A machine learning approach to characterise fabrication porosity effects on the mechanical properties of additively manufactured thermoplastic composites', Journal of Reinforced Plastics and Composites, vol. (In-Press), pp. (In-Press). https://dx.doi.org/10.1177/07316844241236696

DOI 10.1177/07316844241236696 ISSN 0731-6844 ESSN 1530-7964

Publisher: SAGE Publications

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<u>https://creativecommons.org/licenses/by-nc/4.0/</u>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages

(https://us.sagepub.com/en-us/nam/open-access-at-sage)



A machine learning approach to characterise fabrication porosity effects on the mechanical properties of additively manufactured thermoplastic composites

Journal of Reinforced Plastics and Composites 2024, Vol. 0(0) 1–35 © The Author(s) 2024 © UT NG

Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/07316844241236696 journals.sagepub.com/home/jrp



Amadi Gabriel Udu^{1,2}⁽¹⁾, Norman Osa-uwagboe^{2,3}⁽¹⁾, Olusanmi Adeniran⁴⁽¹⁾, Adedeji Aremu⁵, Maryam Ghalati Khaksar¹ and Hongbiao Dong¹

Abstract

The investigation of the mechanical properties of additively manufactured (AM) composite has been the focus of several research over the past decades. However, testing constraints of time and cost have encouraged the exploration of more pragmatic methods such as machine learning (ML) for predicting these characteristics. This study builds on experimental investigations of the flexural, tensile, compressive, porosity, and hardness properties of 3D printed carbon fibre-reinforced polyamide (CF-PA) and carbon fibre-reinforced acrylonitrile butadiene styrene (CF-ABS) composites, proposing the application of ML for predicting these mechanical properties. A comprehensive comparative analysis of various machine learning approaches was executed, with a resultant accuracy ranging between 80 and 99%. The results unveiled the superior predictive performance of ensemble tree learners and the K-NN regressor algorithms when temperature and porosity are selected (based on correlation analysis) as predictors for material hardness and strength in tension, compression, and flexion. In particular, the model built on the extra-tree regressor algorithm demonstrated a remarkably robust fit, with R-squared evaluation scores of 0.9993 and 0.9996 for CF-PA and CF-ABS, respectively. This work develops a ML model that relates porosity to the other mechanical properties of AM composites and the prediction models' exceptional accuracy, along with their precise alignment with experimental data, provide invaluable insights for the autonomous control and data-driven optimization of the structures.

Keywords

Additive manufacturing, damage assessment, machine learning, predictive analysis, mechanical properties

Introduction

Over the past decades, the production and application of AM fibre-reinforced composites have seen a significant surge. This growing interest is predominantly driven by the inherent advantages of AM such as customized design, rapid manufacturing, minimal material wastage, and relatively low costs. When these benefits are paired with the high strengthto-weight ratios of fibre-reinforced composites, the result is a versatile, efficient, and cost-effective solution for various industrial applications.¹ These unique combinations have contributed to the broader adoption and continued evolution of AM fibre-reinforced composites in both research and practical settings.^{2–4} Despite the revolutionary capabilities of AM as a rapid manufacturing technique, the mechanical performance of the fabricated structures depends on critical materials and process factors, such as reinforcement contentment and types, matrix types, and fabrication temperature, among others.^{5–8} For instance, the fabrication temperature affects the degree of porosities, which in turn influences the mechanical performance of the AM fabricated

⁵School of Mechanical Engineering, Coventry University, Coventry, UK

Corresponding author:

Data Availability Statement included at the end of the article

¹School of Engineering, University of Leicester, Leicestershire, UK

 $^{^2\}mbox{Air}$ Force Research and Development Centre, Nigerian Air Force Base, Kaduna, Nigeria

³Wolfson School of Mechanical, Electrical, and Manufacturing Engineering, Loughborough University, Loughborough, UK

⁴Department of Industrial, Manufacturing, and Systems Engineering, Texas Tech University, Lubbock, TX, USA

Norman Osa-uwagboe, Wolfson School of Mechanical, Electrical, and Manufacturing Engineering, Loughborough University, Loughborough LEII 3TU, UK. Email: n.osa-uwagboe@lboro.ac.uk

composites. The assessment of these effects has been primarily through experimental analysis, which can be both time and resource-consuming to determine the extent of the effects. However, recently, the application of machine learning (ML) for predicting and better understanding composite materials performance is gaining traction. This trend can be attributed to ML's high predictive accuracy and robustness across a diverse range of applications.⁹ The use of machine learning in the AM process is showing promising results. Lu et al.¹⁰ utilized ML for real-time defect detection in the fabrication of AM fibre composites. Also, logistics regression was adopted to delineate damaged and undamaged portions of composite structures, leading to the development of highsensitivity damage detection models.^{11–15}

Cai et al.¹⁵ demonstrated ML potential when they used six ML methods to investigate the dynamic strength of AM composite materials. Their results indicated that the artificial neural network (ANN) could achieve the highest prediction accuracy with minimum computational efficiency while the support vector regression (SVR) provided satisfactory prediction with good accuracy and efficiency. The employment of ML in various facets of AM fibre-reinforced composites has been examined by several researchers, each focussing on different aspects of AM composites. Previous work has seen the successful application of different regression models in predicting and optimizing the mechanical properties and fabrication parameters of AM fibre composites. Leon-Becerra et al.¹⁶ utilized the Gaussian process regression (GPR) to describe the principal failure mechanism and predict the tensile stiffness and strength of CF-ABS under various conditions. Zhang et al.¹⁷ developed an ML algorithm to predict the flexural strength of AM-fabricated carbon fibre-reinforced plastic (CFRP) composites, comparing different design factors such as infill patterns, number of reinforcements, and number of concentric carbon rings. Sharma et al.¹⁸ studied the impact of AM process parameters such as wall thickness, print speed, build plate, and extrusion temperatures on the dimensional accuracy of various shapes and compared the results between different materials. In a similar vein, Veeman et al.⁹ used linear regression and ensemble tree learners to optimize process parameters and predict hardness values for Acrylonitrile Butadiene Styrene (ABS) thermoplastic. Furthermore, Vyavahare et al.¹⁹ used deep neural network to predict the strength, stiffness, and specific energy absorption under flexural loading. While also optimizing material properties.

Although considerable progress has been made in the application of ML for predicting and understating various aspects of AM fibre-reinforced composites, existing models often do not sufficiently address the impact of porosity on the overall mechanical performance. Understanding how fabrication temperature-induced porosity affects the performance of these composites is pivotal in developing more reliable and efficient AM processes, ultimately leading to improvement in the mechanical performance of the final product. Motivated by this research gap, our study aims to contribute to the broader

understanding of AM fibre-reinforced composites by incorporating the effects of fabrication temperature-induced porosity into predictive regression models. Thus, the development of an ML framework in this work could provide valuable insights for the optimization and prediction of the mechanical properties of AM fabricated fibre composites. This research is built upon the previous work of the authors, which examined the effects of process and environmental fabrication factors on the tensile, compressive, flexural, and hardness properties of AM fibre-reinforced composites.²⁰⁻²³ We intend to further this investigation by establishing a predictive framework utilizing several ML regression algorithms. Models demonstrating superior performance based on the assessment metrics of mean square error (MSE), mean absolute error (MAE), and coefficient of determination (R^2) were adopted for this research. The selected predictive models were thereafter compared against each other and validated for their accuracy, robustness, and reliability in predicting the mechanical properties of the studied composite. In this research, strength and hardness were chosen as response variables due to their pivotal role in determining the composite's applicability across various industrial applications. The ability of a composite to withstand mechanical stress (strength) and resist deformation (hardness) directly influences its durability and lifespan. On the other hand, temperature induced-porosity was selected as the predictor. Therefore, the application of this methodology in characterizing these properties can aid in designing more efficient and durable AM fabricated fibrecomposites. The remainder of this manuscript is organized as follows: Sections 2 and 3 detail the experimental and ML methodologies, respectively, while the results are discussed in Section 4. In the conclusion, the research findings are encapsulated and their implications for AM fibre-reinforced composites are summarised.

Experimental methodology

For a comprehensive overview of the experimental methodology, the reader is referred to our previous work.^{20–23} These works comprehensively discuss material selection, fabrication processes, sample conditioning, impact setup, dielectric measurement, and the data collection procedure at each predefined measurement interval. In this study, our experimental framework was carried out in adherence to the parameters summarised in Table 1, delineating processing

T	abl	e	Ι.	Material	processing	parameter	s)
---	-----	---	----	----------	------------	-----------	---	---

Parameter	Unit	Value
Infill density	%	100
PET	°C	50 ± 5
Bed temperature	°C	100
Raster angle	degree	0, 90
Layer thickness	mm	0.25
Printing speed	mm/sec	30
Nozzle temperature	°C	230, 250, 270, 290

Test	ASTM standard	Equipment	Test speed	Unit
Tensile	D638	MTS Criterion Model 45	5.0	mm/sec
Rockwell hardness	D785	Clark Tester C12 A	-	-
SEM	-	Thermo Scientific Phenom XL	100X	Magnification
Micro-CT	-	Nikon X-Tex XTH	500	ms

Table 2. Test standard and equipment.



Figure 1. Effect of fabrication temperature on the key properties of AM-fabricated CF-PA composites: (a) porosity volumes, (b) hardness, (c) tensile strength, (d) compressive strength, and (e) flexural strength.



Figure 2. Effect of fabrication temperature on the key properties of AM-fabricated CF-ABS composites: (a) porosity volumes, (b) hardness, (c) tensile strength, (d) compressive strength, and (e) flexural strength.

parameters such as infill density, print speed, and nozzle temperature. To ensure consistency, these parameters were maintained throughout the fabrication of all samples.

In terms of testing, the employed methodologies adhered strictly to ASTM standards and were conducted using equipment, as detailed in Table 2. This adherence to standardized tests assured the reliability and repeatability of the results.

Figures 1 and 2 provide a summary of the experimental test results, specifically the impact of fabrication temperature on the properties of AM-fabricated CF-PA and CF-ABS composites, respectively.



Figure 3. Schematic of the research set-up.

Our comprehensive experimental setup, combined with the implementation of specified processing parameters and standardized tests, enabled the generation of a robust and reliable dataset. This dataset serves as the cornerstone for the development of our machine-learning models, aiming to predict the mechanical properties of AM composite materials. To provide a holistic view of our research methodology, we refer to the comprehensive schematic (Figure 3) included in this section. The schematic illustrates the experimental setup, data collection, and subsequent application of machine learning techniques. It provides a succinct overview of our systematic approach from the initial experiment to the final prediction models, bridging the gap between practical experimentation and data-driven modelling.

Data description

In this study, data was obtained from the porosity, tensile, flexural, compression, and hardness experiments carried out on the two AM composite. Each sample case had a total of five tests. It is worth noting that these mechanical parameters are important to adequately characterize the properties of the materials which could lead to optimization of these structures in engineering applications.^{24–26} Additionally, it has been established that the in-plane and out-of-plane mechanical performance of AM composites are influenced by the porosity gradient which is mainly a product of the manufacturing technique and constituents properties.²¹ Furthermore, no pre-processing technique was employed because it was necessary to capture the underlying patterns in the data distributions as well as avoid influencing the outliers by any pre-processing method. Therefore, the concept adopted in this paper is to develop a predictive model that is robust to outliers.

ML methodology

The machine learning methodology, as outlined in the research schematic (Figure 3) encompasses several fundamental steps that contribute to the development of predictive models. These steps include data normalisation, feature selection, selection of regression algorithms, data resampling, model training, and model evaluation. Each of these components plays an integral role in shaping the accuracy, reliability, and robustness of our ML framework.

Data normalisation

Data normalisation, also known as feature scaling, is an essential pre-processing step in our ML framework. By rescaling the features to a defined range, any potential feature dominance is prevented, the impact of outliers is mitigated, and the convergence and compatibility of the ML algorithms are improved. In this context, the dataset features (represented as $X = [x_1, ..., x_N]^T \in \mathbb{R}^{N \times M}$, where *N* is the total number of observations and M is the number of features)were transformed into a range of 0 to 1. This transformation is performed using

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

where X_{norm} is the normalised feature value, while X_{max} and X_{min} are the maximum and minimum values of the features in the dataset, respectively. This normalisation process ensures that all feature values are equivalently scaled, allowing for a more balanced and effective comparison during the training of the ML models. Other pre-processing steps have been avoided; the experiments were repeated 5 times, to ensure repeatability.

Feature selection

Before proceeding with model training, it is necessary to identify relevant and informative features for accurate prediction. This process, referred to as feature selection, aids in enhancing the model's performance, reducing the risk of overfitting, improving interpretability, and reducing computational complexity.^{27–29} In this study, an integrated approach of expert judgement with Pearson's correlation coefficient (PCC) has been utilised to identify the most informative features. PCC is a statistical measure that assesses the strength and direction of the linear relationship between features.³⁰ It is particularly effective in pinpointing the most informative features that contribute significantly to the target variables. The Pearson's correlation coefficient (r) between the *i*_{th} feature (*x*_i) and the target variable (Y) is given by

$$r(i) = \frac{cov(x_i, Y)}{\sqrt{var(x_i) \cdot var(Y)}}$$
(2)

where cov() and var() are the covariance and variance, respectively. In this work, we identified six predictor features; temperature (°C), porosity volumes (%), strain (mm/ mm), force (*N*), modulus (MPa), and toughness (MJ.m3). These features were selected due to their direct impact on the properties of interest. Meanwhile, the two response variables – the properties we aimed to predict using ML models – were hardness and three types of strength (MPa); tensile, compressive, and flexural. This selection of response variables allows for a comprehensive understanding of the mechanical performance of the materials under study.

ML regressor algorithm selection

Our study utilised Lazy Predict, an intuitive Python library that facilitates efficient training and evaluation of multiple ML models using the default models' configurations and hyperparameters. This tool enabled us to predict the hardness and strength of CF-PA and CF-ABS composites, deploying a pool of 35 ML regression algorithms. Thereafter, a majority voting system was carried out to identify and retain models that achieved a minimum of 70% accuracy in predicting the target variables for the materials investigated in the study. To ascertain that the topperforming algorithms are not a result of chance, the Friedman statistic, a non-parametric statistical test was adopted. It is expressed as

$$\mathcal{X}^{2} = \frac{12}{N(k+1)} \left(\sum R_{j}^{2} - \frac{k(k+1)^{2}}{4} \right)$$
(3)

where N is the number of items, k is the number of treatments, and R^2 is the sum of ranks for the j^{th} treatment.

Owing to its robustness towards assumptions such as data normality, the Friedman test was adopted in this study.³¹ In determining the *p*-value, the computed Friedman statistics is compared to the chi-square distribution. With a *p*-value less than the chosen significance level of 0.05, the



Figure 4. An illustration of regression using Random Forest.

null hypothesis is rejected and thus the regressor is concluded as being of a higher performance and not merely by chance. These selection criteria served to ensure the focus on the most promising, accurate, and reliable models in subsequent analysis and evaluations. Various model predicting algorithms that featured in this study included ensemble learners (e.g. Random Forest), support vector, and k-Nearest Neighbour regressors.

Random forest regressor. Random Forest (RF) is a robust machine learning regressor that utilises ensemble learning (a technique that combines classifiers) to solve regression and classification problems.³² It generates outcomes based on predictions from multiple decision trees and enhances its accuracy of prediction through averaging/majority voting as illustrated in Figure 4.

Support vector regressor. Support vector regressor adopts the principle of support vector machines. Given training data $[(x_1, y_1), ..., (x_n, y_n)]$, SVR seeks a function f(x), that combines weights vector (*w*), input feature (*x*), a bias term (*b*) that defines the hyperplane (i.e. a line that best fits the data points within a certain margin (ε) around the predictor and response values) to make a prediction. Thus, the hyperplane is expressed as

$$f(x) = w^T x + b \tag{4}$$

Subject to the constraints

 $y_i - (w^T x + b) \leq \varepsilon$ for all *i* (upper bound constraint)



Figure 5. Schematic of the support vector regressor.

 $(w^T x + b) - y_i \le \varepsilon$ for all *i* (lower bound constraint)

The objective of SVR is to minimise the L2 regularisation term L2 while satisfying the constraints

$$min: \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \left(\xi_i + \xi_i^*\right)$$
(5)

where *C* is a regularisation parameter that controls the tradeoff between achieving a low error and a large margin; $||w||^2$ is the squared L2 norm of w and while ξ_i and ξ_i^* are slack variables representing prediction deviations from the actual target within the error margin. A schematic representation of SVR is depicted in Figure 5.



Figure 6. Illustration of cross-validation process for one hyperparameter combination.

K-nearest neighbour regressor. The k-Nearest Neighbours (k-NN) regressor is a non-parametric algorithm that predicts continuous values of a target variable by computing the average or weighted average of the k-nearest neighbours in the dataset. It uses the Euclidean distance (or other matrices) to measure the distance between the data point that are to be predicted and all other data points in the training set. It then selects k data points that have the smallest distances to the target data point. Usually, the predicted value for the target data point is the weighted average of the target values of its k-nearest neighbours.

Data resampling and model training

To augment the dataset and enhance the robustness of the analysis, we implemented data resampling techniques, especially bootstrapping integrated with cross-validation. This approach enabled a thorough examination of the model's performance and effective estimation of uncertainties. Bootstrapping is a statistical technique that involves the random selection of *n* samples from the original dataset, with replacement. This preserves the size of the original dataset irrespective of the possibility of duplicate instances being present.³³ By incorporating this technique, we ensured that the training process and the subsequent model evaluation were conducted on the representative set of data, thereby improving the reliability of the ML models.

Hyper-parameter tuning and model training

After selecting the regressor algorithms, hyperparameter tuning is necessary to determine the optimal set of hyperparameters for the respective ML algorithm to improve the model's performance on unseen data. For each of the regressor algorithms, the parameters to be tuned are first identified. These could include the number of estimators, maximum depth of the tree and the number of features to consider when looking for the best fit. In this paper, the Grid Search Cross-Validation (CV) was adopted for hyper-parameter tuning. It creates grid of all possible hyper-parameter combinations with the dataset then partitioned into multiple subsets/folds (usually k = 5 folds), where k-1 folds are used to train the data with the left out used to evaluate the model. The folds are subsequently rotated to ensure that all folds are employed for both training and validation purposes. The performance of the model is then evaluated for each hyperparameter combination by averaging the performance metric across all folds. An illustration of the cross-validation process is presented in Figure 8.

Model evaluation

The effectiveness and accuracy of the models were evaluated using three metrics; the mean squared error (MSE), the mean absolute error (MAE), and the R-squared (R^2) . These metrics have proven to be reliable in assessing the performance of predictive models in regression analysis.^{34–36} MSE measures the average squared difference between the predicted and actual values of a model, while MAE measures the average absolute difference between the true and predicted values of a model. A lower value for both Mean MSE and MAE suggests better model performance. In practical terms, MSE and MAE values closer to 0 are desirable, since they suggest that the model's predictions are very close to the actual values. Due to the squaring operation of MSE, it tends to be more influenced by large errors, making it more sensitive to outliers in the data. In contrast, MAE assigns equal weight to all errors, making it less sensitive to the impact of outliers. Accordingly, for the dataset being utilised in this study, in which the outliers are preserved to capture the underlying patterns, a better MSE value is expected than MAE. The coefficient of determination, \mathbb{R}^2 , quantifies the percentage of variation of v that can be explained by X in the regression model. In practice, R^2 values are typically between 0 and 1, with a value of 1 indicating an excellent predictive performance, as the model can accurately capture all the variability in the dependent variable. The expressions for these metrics are as follows

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$
 (6)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \widehat{y}_i|$$
(7)

$$R^{2} = 1 - \left(\left(\sum_{i=1}^{n} \left(y_{i} - \widehat{y}_{i} \right)^{2} \right) \middle/ \left(\sum_{i=1}^{n} \left(y_{i} - \overline{y} \right)^{2} \right) \right)$$
(8)

where y_i and \hat{y}_i are the true and corresponding predicted values of the response variable Y for the *i*th case, respectively.



Figure 7. Histogram and kde for CF-PA samples: (a) tensile (b) compression, and (c) flexural strength.

Results and discussions

Data analysis

The histogram and kernel density for the CF-PA and CF-ABS material properties are given in Figures 6 and 7, respectively. Based on these results in Figure 7(a), the material properties for tensile strength do not exhibit any clear unimodal patterns, aside from hardness distribution which has a peak in the bin with a lower bound of 0.6 and an upper bound of 0.8, suggesting an unimodal pattern. For porosity, the highest density is in the bin with a lower bound of 0.2 and an upper bound of 0.4, indicating that the material is more likely to have porosity levels within this range. For Figure 8(b), porosity follows a clear

unimodal pattern. Other material properties exhibit more complex distributions with multiple peaks, suggesting variability in the behaviour of the material across different ranges. In Figure 8(c), the distributions for temperature, strain, and porosity appear relatively uniform. The distributions for force, strength, modulus, toughness, and hardness are skewed towards specific ranges, which indicates that these materials could exhibit particular characteristics in those property ranges. The distribution for porosity is, however, slightly skewed towards higher porosity levels, with the highest density (0.5) evident in the bin with a lower bound of 0.8 and an upper bound of 1.0. The distribution is for strain is relatively uniform, with slightly higher density in the bins with lower and upper bounds of 0.4 - 0.6.



Figure 8. Histogram and kde for CF-ABS samples: (a) tensile, (b) compression, and (c) flexural strength.

Set-up

The Python programming language, specifically libraries such as pandas, numpy, sklearn, and matplotlib served as primary tools in developing the ML framework in this study. These libraries facilitated various tasks including numerical computations, preprocessing steps, pipeline initialisation, and model development. For both CF-PA and CF-ABS material samples, the experimental data was loaded using the pandas data frame and normalisation was applied to scale each feature into a range of 0 to 1. Based on the PCC analysis, a general negative correlation was observed with the porosity volume (%) in relation to other features and thus selected as a predictor variable. Additionally, temperature, which exhibited a weak correlation across all material properties, was also chosen as a predictor variable. Accordingly, the temperature and porosity features were indexed as the predictors, while the corresponding hardness and strength (MPa) features were also indexed as the response variables. Next, the Lazy Predict algorithm was initialised to assess model performance on 35 ML regressors. For each material property prediction, the dataset was reshuffled based on the NumPy random state generator and the Lazy Prediction algorithm used for the preliminary model assessment. Thereafter, the Friedman statistics was adopted to assess the significant level of the 10 results computed by Lazy Predict. The results of the Lazy Predict model assessment led to the identification of nine top-performing regressors, based on a majority voting across the material properties to be predicted. A linear regression algorithm was also included as a benchmark algorithm to assess complex nonlinear patterns and interactions between the features.

		I	I

Model	Attribute	Range	Selected value	Mean MSE score
AdaBoost	Learning rate	[0.1, 0.01, 0.001]	0.1	0.0195
	Loss	['linear', 'square', 'exponential']	Exponential	
	n_estimators	[50, 100, 150]	100	
Bagging	Max features	[0.5, 0.8, 1]	I	0.0303
	min_samples	[0.5, 0.8, 1]	I	
	n_estimators	[5, 10, 15, 20]	10	
Decision tree	max_depth	[None, 5, 10]	None	0.0880
	min_samples_leaf	[1, 2, 4]	I	
	min_samples_split	[2, 5, 10]	10	
Extra tree	max_depth	[None, 5,10]	5	0.0283
	max_features	['Auto', 'sqrt', 'log2']	Sqrt	
	min_samples_leaf	[1, 2, 4]	l.	
	min_samples_split	[2, 5, 10]	2	
	n_estimators	[100, 200, 300]	100	
Gradient boosting	Learning rate	[0.1, 0.01, 0.001]	0.01	0.0110
0	max_depth	[3, 5, 7]	3	
	max_features	['Auto', 'sqrt', 'log2']	Sqrt	
	min_samples_leaf	[1, 2, 4]	l.	
	min_samples_split	[2, 5, 10]	10	
	n_estimators	[100, 200, 300]	300	
k-NN	Algorithm	['Brute', 'kd tree', 'ball tree', 'auto']	Auto	0.0149
	leaf_size	[5, 10, 20, 30]	10	
	n_neighbors	[3, 4, 5, 7, 8]	3	
	Weights	['Uniform', 'distance']	Distance	
LightGBM	Learning rate	[1, 0.1, 0.01, 0.001]	I	0.1802
C C	max_depth	[None, I, 3, 5, 7]	None	
	n_estimators	[100, 200, 250, 300]	100	
Random forest	max_depth	[None, 5, 10]	None	0.0365
	min_samples_leaf	[1, 2, 4]	I	
	min_samples_split	[2, 4, 10]	2	
	n_estimators	[100, 200, 300]	300	
SVR	С	[0.001, 0.01, 0.1, 0.5, 1.0]	0.5	0.0330
	Epsilon	[0.1, 0.2, 0.3]	0.1	
	Gamma	['Scale', 'auto']	Scale	
	Kernel	['Linear', 'rbf', 'poly']	rbf	
XGBoost	Learning rate	[1, 0.1, 0.01, 0.001]	0.01	0.0521
	max_depth	[3, 5, 7]	3	
	n_estimators	[100, 200, 300]	200	

Table 3.	Hyper-parameters	of the regressors ar	d their respective investigate	ed ranges for CF-PA in	predicting tensile strength.
----------	------------------	----------------------	--------------------------------	------------------------	------------------------------

Accordingly, the 10 regressor algorithms adopted for developing the prediction models included seven ensemble tree learners (AdaBoost, bagging, decision tree, extra tree, gradient boosting, random forest, and XGBoost regressors), K-nearest neighbour (K-NN) regressor, support vector regressor (SVR), and linear regression. For regressor algorithms in which hyperparameters significantly influenced the model performance., grid search cross-validation was undertaken to determine the optimal hyperparameter values. Each dataset was randomly split into two parts (80% training and 20% testing). Grid search CV was thereafter used to determine the best hyperparameter combination that generalizes well on the regressor algorithm. It is noteworthy that Grid Search CV was undertaken using 80 % training data. The Grid Search was

initialized using k = 5 CV folds with R² set as the scoring parameter. Thereafter, data resampling and model training were conducted, with 20 bootstrap instances. Resampling was done with replacement to maintain the original dataset size. Subsequently, the trained models were evaluated using the MSE, MAE, and R² values. The actual versus the predicted values were plotted, aiding in more intuitive understanding of the model's accuracy.

Performance

PCC analysis. Figures 9 and 10 present the PCC for (a) tensile, (b) compression, and (c) flexural material properties of CF-PA and CF-ABS, respectively. For clarity, the features have been



Figure 9. PCC for material properties of CF-PA: (a) tensile, (b) compression, and (c) flexural.

abbreviated as follows: temperature – Tc, porosity volume – P, force – F, strain – Sm, toughness – Tm, modulus – M, strength – Sp and, hardness – H. Based on the PCC analysis, a generally weak correlation exists between temperature other tensile and compression material properties of CF-ABS. This correlation pattern was similarly observed in compression material properties of CF-PA. Aside from the case of compression for CF-PA (Figure 9(a)), porosity generally exhibited a weak negative correlation with other features for all the samples tested (Figures 9(b) and (c), Figures 10(a–c)). The modulus feature also presented a negative correlation for the flexural material property for CF-ABS. Interestingly, a strong correlation was also observed between strain and toughness across all material properties, underscoring their interconnected roles in material performance.

Friedman statistic results. For all the material properties prediction considered, a p-value >0.05 was realised after the second iteration, which suggested that the Lazy

Predict algorithm and the majority voting system adopted in selecting the top performing model was not a result of chance. Figure 11 is a plot of the Friedman Statistic and *p*-value results for CF-PA, hardness prediction based on 10 iterations. The Friedman Statistic and *p*-value at the second iteration was 66.41 and 0.0054, respectively.

Hyper-parameter tuning results. Table 3 shows the hyperparameters, ranges, selected parameters, and mean MSE test score for CF-PA in predicting the tensile strength based on temperature and porosity. Other hyperparameters not shown are at their default state.

Following this, the model underwent training using the training data, and predictions were generated for both the training and testing datasets. Subsequently, the R², MSE, and MAE values were calculated, accompanied by plots depicting the actual and predicted values for each respective model.



Figure 10. PCC for material properties of CF-ABS (a) tensile, (b) compression, and (c) flexural.

Evaluation results. The inclusion of tree learners and K-NN among the selected regressor algorithms considered in the study demonstrated their ability to capture non-linear relationships and interactions between features as well as their robustness in handling outliers. An added advantage of these algorithms is that they do not make any assumptions about the underlying data distribution, an attribute that is particularly beneficial for datasets where the distribution may not be readily modelled. Notably, the low performance of the linear regression model can likely be attributed to its limitation in dealing with non-linear patterns and relationships. This underlines the need to consider more flexible, non-linear models when dealing with multidimensional and intricate relationships among features. Tables 4-7 are the model evaluation results of hardness, tensile, compressive, and flexural material property prediction for CF-PA and CF-ABS, respectively (with the best performing models emphasized in the respective tables).

The models with the highest performance for predictions, based on the MSE, MAE, and R^2 values, are highlighted in this section. Excluding the linear regression model, which was added as a benchmark, all models delivered a predictive performance between 80 % and 99 %. Owing to the sensitivity of MSE to outliers, it delivered better performance results than the MAE values, for all material properties prediction considered. This is presented in Figures 12 and 13, which are the bar charts for the best-performing models for CF-PA and CF-ABS, respectively.

The models built on k-NN provided the optimal performance for predicting hardness and flexural strength properties for both CF-PA and CF-ABS composites. For predicting compressive strength, the models built on decision trees and extra tree regressors delivered peak R²



Figure 11. Friedman Statistic and p-value results for CF-PA - hardness based on 10 iterations.



Figure 12. Best performing models for CF-PA; (a) hardness, (b) tensile, (c) compression, and (d) flexural.

performances of 0.9885 for CF-PA and 0.9566 for CF-ABS, respectively. The highest accuracy was observed in the prediction of tensile strength, yielding an MSE, MAE, and R^2 of 0.0001, 0.0024, and 0.9996 for CF-PA, and 0.0001, 0.0036, and 0.9993 for CF-ABS. The graphical representation of the actual versus predicted values for the highest-performing models for hardness, tensile, compressive, and flexural strengths for both CF-PA and CF-ABS are provided in Figures 14 and 15 respectively. Additional plots for the remaining models are appended in Appendix 1.

The presented plots illustrate the relationship between the actual and predicted values in relation to the ideal fit line. Specifically, the actual and predicted values for the tensile strength for CF-PA and CF-ABS, as displayed in Figures 14(b) and 15(b) respectively, are more closely aligned on the perfect fit line than any other models. Additional lines have been included in the plots to account for the deviations from the actual predictions. Dotted lines represent a range of ± 0.1 deviations from the actual values and can be interpreted as a reasonable approximation to the actual values. The dashed lines, indicating a ± 0.2 deviation from the actual values, represent a slightly larger, but still acceptable deviation. Points falling outside of this ± 0.2 range represent



Figure 13. Best performing models for CF-ABS; (a) hardness, (b) tensile, (c) compression, and (d) flexural.



Figure 14. Actual versus predicted values of optimal performing models for CF-PA: (a) hardness (K-NN), (b) tensile (Extra Tree), (c) compressive (Decision Tree), and (d) flexural strength (K-NN).



Figure 15. Actual versus predicted values of optimal performing models for CF-ABS; (a) hardness (K-NN), (b) tensile (Extra Tree), (c) compressive (Extra Tree), and (d) flexural strength (K-NN).

	CF-PA			CF-ABS	CF-ABS	
Models	MSE	MAE	R ²	MSE	MAE	R ²
AdaBoost	0.0052	0.0511	0.9320	0.0060	0.0315	0.9370
Bagging	0.0076	0.0629	0.9009	0.0082	0.0567	0.9148
Decision tree	0.0059	0.0379	0.9235	0.0113	0.0364	0.8822
Extra tree	0.0050	0.0327	0.9348	0.0059	0.0223	0.9384
Gradient boosting	0.0059	0.0406	0.9224	0.0057	0.0219	0.9411
K-NN	0.0025	0.0235	0.9673	0.0032	0.0205	0.9664
Linear regression	0.0140	0.1003	0.8166	0.0237	0.1338	0.7529
Random forest	0.0064	0.0624	0.9169	0.0060	0.0616	0.9379
SVR	0.0106	0.0932	0.8610	0.0113	0.0962	0.8826
XGBoost	0.0058	0.0384	0.9242	0.0112	0.0369	0.8833

Table 4. Model evaluation results for hardness material property prediction.

The bold values in the tables used to highlight the best performing models.

significant prediction errors. Among the top-performing models, only Figure 14(d), which predicts the CF-PA flexural strength prediction, has a predicted value falling outside this range, accounting for the R^2 value of 0.9353. In Figure 14(a), which shows the hardness prediction for

CF-PA, three actual values were falsely predicted at various points (from 0.796 to 0.818). Hence, the model can guarantee a prediction of up to 99 % accuracy except for hardness properties with fabrication temperature-induced porosity volume within the range of 19.3

	CF-PA			CF-ABS				
Models	MSE	MAE	R ²	MSE	MAE	R ²		
AdaBoost	0.0094	0.0460	0.9064	0.0006	0.0139	0.9950		
Bagging	0.0085	0.0618	0.9316	0.0056	0.0475	0.9555		
Decision tree	0.0003	0.0064	0.9977	0.0003	0.0064	0.9977		
Extra tree	0.0001	0.0024	0.9996	0.0001	0.0036	0.9993		
Gradient boosting	0.0003	0.0065	0.9977	0.0003	0.0065	0.9977		
K-NN	0.0003	0.0062	0.9975	0.0020	0.0146	0.9837		
Linear regression	0.0207	0.1096	0.8339	0.0207	0.1096	0.8339		
Random forest	0.0038	0.0427	0.9695	0.0055	0.0490	0.9558		
SVR	0.0142	0.1056	0.8589	0.0098	0.0860	0.9215		
XGBoost	0.0003	0.0070	0.9978	0.0003	0.0070	0.9978		

Table 5. Model evaluation results for tensile strength material property prediction.

The bold values in the tables used to highlight the best performing models.

Table 6. Model evaluation results for compressive strength material property prediction.

	CF-PA			CF-ABS		
Models	MSE	MAE	R ²	MSE	MAE	R ²
AdaBoost	0.0027	0.0276	0.9761	0.0079	0.0362	0.9126
Bagging	0.0139	0.0927	0.8778	0.0134	0.0855	0.8515
Decision tree	0.0013	0.0132	0.9885	0.0114	0.0324	0.8741
Extra tree	0.0026	0.0197	0.9796	0.0039	0.0243	0.9566
Gradient boosting	0.0013	0.0133	0.9884	0.0077	0.0284	0.9144
K-NN	0.0073	0.0327	0.9361	0.0060	0.0247	0.9340
Linear regression	0.0576	0.1765	0.4943	0.0287	0.1426	0.6819
Random forest	0.0058	0.0558	0.9490	0.0113	0.0812	0.8747
SVR	0.0102	0.0812	0.9106	0.0148	0.1131	0.8359
XGBoost	0.0013	0.0139	0.9882	0.0071	0.0252	0.9208

The bold values in the tables used to highlight the best performing models.

Table 7. Model evaluation results for flexural strength material property prediction.

	CF-PA			CF-ABS			
Models	MSE	MAE R ²		MSE	MAE	R ²	
AdaBoost	0.0088	0.0408	0.8974	0.0041	0.0357	0.9702	
Bagging	0.0108	0.0823	0.8734	0.0083	0.0707	0.9391	
Decision tree	0.0082	0.0281	0.9043	0.0036	0.0246	0.9733	
Extra tree	0.0082	0.0283	0.9041	0.0047	0.0245	0.9656	
Gradient boosting	0.0082	0.0299	0.9039	0.0042	0.0248	0.9693	
K-NN	0.0055	0.0245	0.9353	0.0036	0.0244	0.9736	
Linear regression	0.0256	0.1104	0.7002	0.0417	0.1830	0.6933	
Random forest	0.0107	0.0680	0.8744	0.0074	0.0683	0.9457	
SVR	0.0236	0.1108	0.7238	0.0100	0.0861	0.9267	
XGBoost	0.0082	0.0289	0.9042	0.0042	0.0250	0.9690	

The bold values in the tables used to highlight the best performing models.

%-19.9 % which corresponds to normalized predictability values of 0.796 to 0.818, respectively.

Conclusion

This study employed a ML approach to characterise the mechanical properties of AM fabricated composites. The findings reveal a generally negative correlation between porosity volume (%) and the tensile and compressive strength for both CF-PA and CF-ABS structures. A variety of ML regressor algorithms were explored in predicting the materials' properties including hardness and strength (tensile, compressive, and flexural). In particular, the study explored the impact of fabrication temperature-induced porosity on selected mechanical properties of AM CF-PA and CF-ABS composite structures. The prediction models' results revealed that ensemble tree learners and the K-NN regressor algorithms delivered the most accurate results when temperature-induced-porosity was selected as predictors for material hardness and strength. This demonstrated the ability of the ML algorithms to capture non-linear relationships and interactions, as well as robustness to outliers, in contrast to the lower performance of the linear regression model which was benchmarked for the study. The presented models achieved an accuracy of between 80% and 99%. The high performance of the test model can be applied to overcome the reliance on experimental and destructive techniques, as well as compensate for limitations in the technical skills of AM equipment operators in characterizing the mechanical properties and damage phenomenon of CF-PA and CF-ABS AM composite structures. Notably, the model built on the extra tree regressor algorithm delivered the highest evaluation scores for both CF-PA and CF-ABS, with R² values of 0.9993 and 0.9996 respectively. In summary, the developed model can guarantee an accurate prediction for all samples except for CF-PA with fabrication temperature-induced porosity volume within the range of 19.3 %-19.9 %. Furthermore, this model would be valuable in enhancing control and providing autonomy while alleviating time and cost constraints in both research and industrial applications of AM structures.

Declaration of conflicting interest

The author(s) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Petroleum Technology Development Fund under grant PTDF/ED./OSS/PHD/AGU/1076/17.

ORCID iDs

Amadi Gabriel Udu b https://orcid.org/0000-0001-8944-4940 Norman Osa-uwagboe b https://orcid.org/0000-0002-6560-445X Olusanmi Adeniran b https://orcid.org/0000-0001-9908-7246

Data Availability Statement

The data given in this article are available on request.

References

- Adil S and Lazoglu I. A review on additive manufacturing of carbon fiber-reinforced polymers: current methods, materials, mechanical properties, applications and challenges. *J Appl Polym Sci* 2023; 140: e53476, Epub ahead of print 15 February 2023. DOI: 10.1002/app.53476.
- Angrish A. A critical analysis of additive manufacturing technologies for aerospace applications. In: *IEEE Aerospace Conference Proceedings*. Washington: IEEE Computer Society, 2014. Epub ahead of print 2014. DOI: 10.1109/AERO. 2014.6836456.
- Pervaiz M, Panthapulakkal S, Kc B, et al. Emerging trends in automotive lightweighting through novel composite materials. *Mater Sci Appl* 2016; 07: 26–38.
- Goh GD and Yeong WY. Mode I interlaminar fracture toughness of additively manufactured carbon fibre thermoplastic. In: *Proceedings of the International Conference on Progress in Additive Manufacturing*. Detroit, Michigan: Pro-AM, 2018, pp. 505–510.
- Ibrahim Y, Melenka GW and Kempers R. Additive manufacturing of continuous wire polymer composites. *Manuf Lett* 2018; 16: 49–51.
- Adeniran O, Cong W and Aremu A. A review of material design factors in the additive manufacturing of short carbon fiber reinforced plastic composites. In: *Advances in Industrial and Manufacturing Engineering*. Berlin, Germany: Springer Nature; 2022, 100100.
- Ivey M, Melenka GW, Carey JP, et al. Characterizing shortfiber-reinforced composites produced using additive manufacturing. Adv Manuf Polym Compos Sci 2017; 3: 81–91.
- Parsazadeh M, Sharma S and Dahotre N. Towards the next generation of machine learning models in additive manufacturing: a review of process dependent material evolution. *Prog Mater Sci* 2023; 135: 101102, Epub ahead of print 1 June 2023. DOI: 10.1016/j.pmatsci.2023.101102.
- Veeman D, Sudharsan S, Surendhar GJ, et al. Machine learning model for predicting the hardness of additively manufactured acrylonitrile butadiene styrene. *Mater Today Commun* 2023; 35: 106147, Epub ahead of print 1 June 2023. DOI: 10.1016/j.mtcomm.2023.106147.
- Lu L, Hou J, Yuan S, et al. Deep learning-assisted real-time defect detection and closed-loop adjustment for additive manufacturing of continuous fiber-reinforced polymer composites. *Robot Comput Integr Manuf* 2022; 79: 102431, Epub ahead of print 1 February 2022. DOI: 10.1016/j.rcim.2022. 102431.
- Idolor O, Berkowitz K, Guha RD, et al. Nondestructive examination of polymer composites by analysis of polymerwater interactions and damage-dependent hysteresis. *Compos Struct* 2022; 287: 115377.

- Idolor O, Guha RD, Berkowitz K, et al. An experimental study of the dynamic molecular state of transient moisture in damaged polymer composites. *Polym Compos* 2021; 42: 3391–3403.
- 13. Gljušćić M, Lanc D, Franulović M, et al. Microstructural analysis of the transverse and shear behavior of additively manufactured CFRP composite RVEs based on the phase-field fracture theory. *Journal of Composites Science* 2023; 7: 38, Epub ahead of print 1 January 2023. DOI: 10.3390/jcs7010038
- Yang B, Fu K, Lee J, et al. Artificial neural network (ANN)-Based residual strength prediction of carbon fibre reinforced composites (CFRCs) after impact. *Appl Compos Mater* 2021; 28: 809–833.
- Cai R, Wang K, Wen W, et al. Application of machine learning methods on dynamic strength analysis for additive manufactured polypropylene-based composites. *Polym Test* 2022; 110: 107580, Epub ahead of print 1 June 2022. DOI: 10.1016/ j.polymertesting.2022.107580.
- Leon-Becerra J, González-Estrada OA and Sánchez-Acevedo H. Comparison of models to predict mechanical properties of FR-AM composites and a fractographical study. *Polymers* 2022; 14: 3546, Epub ahead of print 1 September 2022. DOI: 10.3390/polym14173546.
- Zhang Z, Shi J, Yu T, et al. Predicting flexural strength of additively manufactured continuous carbon fiber- reinforced polymer composites using machine learning. *J Comput Inf Sci Eng* 2020; 20: 061015, Epub ahead of print 1 December 2020. DOI: 10.1115/1.4047477.
- Sharma P, Vaid H, Vajpeyi R, et al. Predicting the dimensional variation of geometries produced through FDM 3D printing employing supervised machine learning. *Sensors International* 2022; 3: 100194,Epub ahead of print 1 January 2022. DOI: 10.1016/j.sintl.2022.100194.
- Vyavahare S, Teraiya S and Kumar S. FDM manufactured auxetic structures: an investigation of mechanical properties using machine learning techniques. *Int J Solids Struct* 2023; 265–266: 112126.
- Adeniran O, Cong W, Bediako E, et al. Environmental affected mechanical performance of additively manufactured carbon fiber–reinforced plastic composites. *J Compos Mater* 2022; 56: 1139–1150.
- Adeniran O, Osa-uwagboe N, Cong W, et al. Fabrication temperature-related porosity effects on the mechanical properties of additively manufactured CFRP composites. *J. Compos. Sci* 2023; 7: 1–15.
- Adeniran O, Cong W and Oluwabunmi K. Thermoplastic matrix material influences on the mechanical performance of additively manufactured carbon-fiber-reinforced plastic composites. *J Compos Mater* 2022; 56: 1391–1405.
- 23. Adeniran O, Cong W, Bediako E, et al. Additive manufacturing of carbon fiber reinforced plastic composites: the effect of fiber content on compressive properties. *Journal* of Composites Science 2021; 5: 325, Epub ahead of print 1 November 2021. DOI: 10.3390/jcs5120325.
- 24. Lee D, Kwon HJ, Yang S, et al. Fabrication, testing, and analysis of sandwich structure with composite skin and

additive manufactured core. *J Reinforc Plast Compos* 2021; 40: 654–664.

- Hofstätter T, Pedersen DB, Tosello G, et al. State-of-the-art of fiber-reinforced polymers in additive manufacturing technologies. *J Reinforc Plast Compos* 2017; 36: 1061–1073.
- 26. Akhoundi B. An evaluation of the shape-memory behavior and mechanical properties of polylactic acid/ Ni80Cr20 continuous wire composite produced by extrusion-based additive manufacturing and in-melt simultaneous impregnation method. J Reinforc Plast Compos 2023; 1–15, Epub ahead of print 2023. DOI: 10. 1177/07316844231197036
- Effrosynidis D and Arampatzis A. An evaluation of feature selection methods for environmental data. *Ecol Inform* 2021; 61: 101224.
- Cai J, Luo J, Wang S, et al. Feature selection in machine learning: a new perspective. *Neurocomputing* 2018; 300: 70–79.
- Dhal P and Azad C. A comprehensive survey on feature selection in the various fields of machine learning. Guildford: Applied Intelligence, 2022. Epub ahead of print 2022. DOI: 10.1007/s10489-021-02550-9.
- Chandrashekar G and Sahin F. A survey on feature selection methods. *Comput Electr Eng* 2014; 40: 16–28.
- López-Vázquez C and Hochsztain E. Extended and updated tables for the Friedman rank test. *Commun Stat Theor Methods* 2019; 48: 268–281.
- 32. Sagi O and Rokach L. Ensemble learning: A survey. *Wiley* Interdiscip Rev Data Min Knowl Discov 2018; 8: 1–18.
- Tsamardinos I, Greasidou E and Borboudakis G. Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation. *Mach Learn* 2018; 107: 1895–1922.
- Montgomery DC. Applied statistics and probability for engineers. 6th ed. New Jersey: John Wiley & Sons Limited, 2014.
- Botchkarev A. A new typology design of performance metrics to measure errors in machine learning regression algorithms. *Interdiscipl J Inf Knowl Manag* 2019; 14: 45–76.
- 36. Chicco D, Warrens MJ and Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci* 2021; 7: 1–24.

Appendix

Appendix I

The actual and predicted values of CF-PA hardness, tensile compressive and flexural strength based on 10 regressor models considered in the study are depicted in Figures A1–A4 respectively, while those for CF-ABS are presented in Figures A5–A8 respectively.



Figure AI. Continued.

Figure A1. (continued)



Figure A1. Actual and predicted values of CF-PA hardness.



Figure A2. Continued.

Figure A2. (continued)



Figure A2. Actual and predicted values of CF-PA tensile strength.



Figure A3. Continued.

Figure A3. (continued)



Figure A3. Actual and predicted values of CF-PA compressive strength.



Figure A4. Continued.

Figure A4. (continued)



Figure A4. Actual and predicted values of CF-PA flexural strength.



Figure A5. Continued.

Figure A5. (continued)

Figure A5. Actual and predicted values of CF-ABS hardness.

Figure A6. Continued.

Figure A6. (continued)

Figure A6. Actual and predicted values of CF-ABS tensile strength.

Figure A7. Continued.

Figure A7. (continued)

Figure A7. Actual and predicted values of CF-ABS compressive strength.

Figure A8. Continued.

Figure A8. (continued)

Figure A8. Actual and predicted values of CF-ABS flexural strength.