

Article



Prediction of Fracture Toughness of Pultruded Composites Based on Supervised Machine Learning

Radmir Karamov * D, Iskander Akhatov D and Ivan V. Sergeichev D

Center for Materials Technologies, Skolkovo Institute of Science and Technology, Bolshoy Boulevard 30, bld. 1, 121205 Moscow, Russia

* Correspondence: radmir.karamov@skoltech.ru

Abstract: Prediction of mechanical properties is an essential part of material design. State-of-the-art simulation-based prediction requires data on microstructure and inter-component interactions of material. However, due to high costs and time limitations, such parameters, which are especially required for the simulation of advanced properties, are not always available. This paper proposes a data-driven approach to predicting the labor-consuming fracture toughness based on a series of standard, easy-to-measure mechanical characteristics. Three supervised machine-learning (ML) models (artificial neural networks, a random forest algorithm, and gradient boosting) were designed and tested for the prediction of mechanical properties of pultruded composites. A considerable dataset of mechanical properties was acquired as results of standard tensile, compression, flexure, inplane shear, and Charpy tests and utilized as the input to predict the fracture toughness. Furthermore, this study investigated the correlations between the obtained mechanical characteristics. Analysis of ML performance showed that fracture toughness had the highest correlations with longitudinal bending and transverse tension and a strong correlation with the longitudinal compression modulus and tensile strength. The gradient boosting decision tree-based algorithms demonstrated the best prediction performance for fracture toughness, with an MSE less than 10% of the average value, providing a prediction within the range of experimental error. The ML algorithms showed potential in terms of determining which macro-level parameters can be used to predict micro-level material characteristics and how. The results provide inspiration for future pultruded composite material design and can enhance the numerical simulations of material.

Keywords: composite materials; pultrusion; fracture toughness; machine learning; correlation

1. Introduction

Prediction of the mechanical properties of novel composite materials is one of the primary goals of research in the field of material design. The prediction of material properties is nowadays attempted by using physics-based simulations: molecular dynamics, finite element methods (FEMs), and others [1–4]. However, these methods have some limitations due to the unknown microstructure of materials, computational expensiveness, etc. [5,6]. In addition, a propagation error can be introduced in the multiscale modeling of large systems [7,8]. Another way to predict the mechanical properties of materials is the data-driven approach, which has recently become more popular and has given us some inspiration for a novel methodology for the design and characterization of composite materials.

Data-driven approaches based on machine-learning algorithms have been applied in material science in recent decades, accelerating the design and discovery of new functional and structural materials [9]. Machine learning (ML) is a branch of artificial intelligence that allows us to analyze large, noisy datasets and learn and detect data patterns and correlations between input and output variables by optimizing the chosen machine-learning model [10]. This approach is mainly used in the composite materials field to design and discover new materials and their basic properties, such as stiffness and strength.



Citation: Karamov, R.; Akhatov, I.; Sergeichev, I.V. Prediction of Fracture Toughness of Pultruded Composites Based on Supervised Machine Learning. *Polymers* **2022**, *14*, 3619. https://doi.org/10.3390/ polym14173619

Academic Editors: Mikhail G. Kiselev and Yury Budkov

Received: 4 August 2022 Accepted: 30 August 2022 Published: 1 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). One of the first works in this field by Mukherjee et al. [11] proposed to predict the yield stress under the tension of metal matrix composites using artificial neural networks with microstructure factors, such as the volume fraction, fiber arrangements, and the properties of the components, as input variables. However, they employed a database that was synthetically generated using the FEM approach to acquire the necessary amount of data due to the insignificant progress in the training algorithms for such models at the time. The authors of [12,13] proposed a methodology to predict composites' mechanical properties using a very small-sized database. Nevertheless, this work was criticized [14]: the neural networks in their work were too complicated for the given dataset, and highly correlated input and output variables were also used.

In recent years, researchers have worked on predicting the mechanical properties of composite materials based on their inner structure using more advanced machinelearning techniques [15–22]. For instance, the authors of [23] recently applied linear and convolutional neural network (CNN) models to predict the toughness and strength of 2D functional composite systems based on images of their microstructure. They performed calculations using TensorFlow [24], a general-purpose ML framework that has the ability to search for optimal designs with limited information. Yang et al. [25] demonstrated the implementation of a deep-learning, feature-engineering-free approach for predicting the microscale elastic strain field in a given 3D voxel-based microstructure of a high-contrast, two-phase composite. The results showed that deep learning approaches could implicitly learn salient information about local neighborhood details. However, there are still only a limited number of studies that investigate the ML prediction of the advanced mechanical properties of composite materials based on standard, easy-to-measure properties and that discuss their ML correlations without considering the material's inner structure in detail.

More work on data-driven approaches can be found for other materials but, unfortunately, the published models cannot be applied to predict the mechanical properties of composite materials due to different output parameters and larger datasets. For example, Tiryaki and Aydın [26] applied this methodology to design an artificial neural network model to predict the compression strength of heat-treated wood without comprehensive experiments. The results indicated that the artificial neural network model provided a better prediction than the multiple linear regression model. The strength properties of the heat-treated wood could be determined in a short time with low error rates, allowing the usability of such wood species for structural purposes to be better understood.

One of the most compelling studies [27] was conducted on engineering alloys and described a methodology to acquire elastic, and even plastic, properties based on one sensitive test (instrumented indentation) and the latest developments in deep learning and neural networks. However, this methodology did not consider other machine-learning algorithms, and it is only applicable to instrumented indentation of homogeneous materials and cannot be applied for heterogeneous composite materials. Additionally, the machine-learning correlations between fracture toughness and other material properties were not investigated.

In this study, we propose a new way to predict fracture toughness and to determine its correlations with other mechanical properties based on machine-learning algorithms. We designed and analyzed ML methodologies to predict the mechanical properties of pultruded composites and their correlations throughout the length of the conditionally infinite profile. Pultrusion was selected as it is one of the continuous manufacturing techniques that can provide high-quality and cost-effective composite materials. To characterize a pultruded material with high accuracy, it is necessary to perform a large number of mechanical tests along the produced structural profile, which is a time-consuming and costly task [28,29]. Despite comprehensive studies of pultruded composites performed over the last thirty years (recently reviewed in [30]), there still exists a lack of knowledge about the correlations between different material properties as they are found in the structure. This paper consists of the following parts: Materials and Methods and Results and Discussions. In the following section, the materials and methods used are thoroughly described. In Section 3, the fracture toughness prediction results are presented and the ML-based correlation with easy-to-measure characteristics is discussed.

2. Materials and Methods

2.1. Pultruded Composite Material

The material used to illustrate and analyze the machine-learning prediction of the fracture toughness was pultruded glass fiber reinforced polymer (GFRP). The material (Figure 1) was manufactured using a Pultrex $P500 \times 6T$ with a pulling speed of 0.4 m/min and a temperature of 125 °C.



Figure 1. Pultruded profile side cross-section (side view).

The material was produced in two days; it was possible to observe deviations in the batches from different days due to different levels of temperature and humidity and human factors.

2.2. Mechanical Testing

A 50 m pultruded profile was produced. From each meter, ten specimens for the standard mechanical tests and three specimens for the fracture toughness tests were cut and tested (Figure 2). This allowed us to create a dataset of 50 batches with different properties from along the length of the material that could be used for ML training.



Figure 2. Cutting scheme for specimens from one meter of the pultruded profile (the specimen numbers are indicated for internal use).

2.2.1. Standard Mechanical Tests

Several standard mechanical properties were chosen to analyze the machine-learning performance with possible correlations with the fracture toughness. The material's mechanical properties were obtained during a mechanical characterization study consisting of tension, compression, flexure, in-plane shear, and Charpy impact tests. All tests were undertaken in 0° and 90° directions relative to the pulling direction.

The test standards, testing machine, and obtained properties are given in Table 1.

Mechanical Test	Standard	Machine	Strain Measurements	Output Properties
Tension	ASTM D3039			Tensile modulus, strength
Compression	ASTM D6641 CLC	Instron 5969	DIC	Compression modulus, strength
Flexure	ASTM D7264			Flexure modulus, strength
In-plane shear	ASTM D7078			Shear strength, shear modulus
Charpy impact	ASTM D256	Instron CEAST 9340	Strain gauge 64k	Failure energy

Table 1. Summary of mechanical tests.

2.2.2. Fracture Toughness Test

The fracture toughness behavior of pultruded glass fiber reinforced materials has only been described in a few experimental studies [31,32]. In this work, we adopted a wide compact tension method to characterize the transverse fracture properties of the material. This method was proposed by Almeida-Fernandes et al. [33]. The method was proved to be "effective in achieving a stable propagation stage", and the final properties showed "good agreement across visually based methods". The test involves applying tensile 20 N/mm loadings to a specially prepared specimen (Figure 3) and observing crack growth. The energy release rate G_{Ic}^{lam} and the stress intensity factor in mode I of loading K_{Ic} (plane strain fracture toughness) were estimated using the following formulae [34]:

$$G_{Ic}^{lam} = \frac{K_{Ic}^2}{\sqrt{2E_{11}E_{22}}} \sqrt{\sqrt{\frac{E_{11}}{E_{22}}} + \frac{E_{11}}{2G_{12}}} - v_{12},$$
(1)

$$K_{Ic} = \frac{P}{t\sqrt{w}} f\left(\frac{a}{w}\right),\tag{2}$$

$$f\left(\frac{a}{w}\right) = \frac{2 + \frac{a}{w}}{\left(1 - \frac{a}{w}\right)^{1.5}} \left[0.886 + 4.64 \left(\frac{a}{w}\right) - 13.32 \left(\frac{a}{w}\right)^2 + 14.72 \left(\frac{a}{w}\right)^3 - 5.6 \left(\frac{a}{w}\right)^4 \right]$$
(3)

where E_{11} and E_{22} are the elastic moduli in the crack (longitudinal) and the load (transverse) directions, respectively; G_{12} is the shear elastic modulus; v_{12} is the Poisson's ratio; P is the load at a given crack length a; t is the specimen thickness; and w is the geometric parameter of the specimen (see Figure 3).



Figure 3. Specimen geometry for the fracture toughness test.

Fracture toughness experiments take up to ten times longer to complete compared to standard mechanical tests. More detailed information about fracture toughness tests for pultruded GFRP is presented in [33].

We used a stress intensity factor K_{Ic} as a fracture toughness characterization because the energy release rate is highly dependent on the other characteristics.

2.3. Machine-Learning Methods

Prediction of one property from others is a regression problem, and the use of machinelearning (ML) algorithms can be presented as a parametric function:

$$y = f(x), \tag{4}$$

where *x* and *y* are vectors for the input and output (to be predicted) data, respectively. ML techniques establish which parameters are used in the function and how they are calculated from the training dataset. The training dataset is represented by input–output vector pairs: $(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)$.

Many ML techniques exist, and, the most popular regression models were implemented in this study; namely, artificial neural networks, a random forest algorithm, and gradient boosting decision trees. We also tried support vector regression and Gaussian process regression but, in this study, they tended to output mean values and were not included in the paper.

To estimate the ML models' performances, cross-validation was performed: the models were trained with 90% of the data as training data and 10% as test data. The cross-validation was performed 50 times with different training and test data to acquire statistics for the ML models' performances.

2.3.1. Artificial Neural Network

An artificial neural network usually consists of three types of layers: input data, hidden layers, and output data (Figure 4).



Figure 4. Example of fully connected neural network with two input neurons (input layer), nine hidden neurons (two hidden layers), and one output neuron (output layer).

There is always one input and one output layer in a neural network, but the number of hidden layers may vary. Each layer includes neurons that are described by their values. Usually, fully connected neural networks are used for regression problems, where each neuron in the hidden and output layers is connected to all neurons of the previous layer. The values for each neuron are calculated by summing up the values for all neurons from previous layers, multiplied by weights, and adding biases, which can be thought of as analogous to a constant shift in a function. This logic is represented by a set of straight lines in Figure 4. To introduce additional nonlinearity to the algorithm, some nonlinear activation functions (denoted as σ) were applied to each neuron as the overall weight [35]. The mathematical representation of such an algorithm for the calculation of the neuron values $a_{(n+1)}$ of layer n + 1 is the following:

$$\boldsymbol{a}_{(n+1)} = \sigma \Big(\boldsymbol{W} \boldsymbol{a}_{(n)} + \boldsymbol{b} \Big).$$
⁽⁵⁾

The parameters we needed to find using the dataset of the input and output values were W weights and b biases. One of the approaches to do that is the backpropagation optimization method, which iteratively updates weights and biases using gradient descent-

based algorithms, minimizing the error between the predicted values and the values from the dataset [36].

In this study, a fully connected neural network was used with mean squared error loss. Several NN architectures were tested, the architecture with the best performance consisted of 19 neurons in the input layer; three hidden layers with 20, 40, and 10 neurons; and 1 neuron in the output layer. After selecting seven main features, seven input neurons were used with three hidden layers (7, 14, and 4 neurons), and one neuron in the output layer. An Adam optimization method was employed [37]. The learning rate was 0.01, and the number of epochs was 3000. The model was implemented in the TensorFlow 2.5 framework.

2.3.2. Random Forest

The random forest algorithm is an advanced ensemble algorithm [38]. The fundamental elements of this algorithm are decision trees (Figure 5). For regression problems, a decision tree can be represented as a function created by recursively partitioning each independent variable. Based on the partition, the output value is predicted. The partitioning is performed to optimize the mean squared errors (MSEs) between the predicted and actual values. The optimal tree is the smallest tree that has the minimum relative error with cross-validation.



Figure 5. Example of one regression decision tree with depth equal to 4.

The random forest regression algorithm consists of N (usually more than 100) independent decision trees with a random subset of independent variables. To render the decision trees in the random forest algorithm independent, the bagging method [39] was used. A machine-learning model with robust resistance to overfitting was created by combining a random subset of input variables and random training-set selection. The final prediction of a random forest model is an unweighted average of the predictions of all the individual trees.

Although random forest models cannot be graphically presented as decision trees, the variable importance measures (VIMs) can be calculated [40,41]. In the VIM concept, the impact of optimizing an input variable partition is measured with respect to the change in the MSE. The greater the prediction accuracy reduction is during optimization, the more significant the impact of this variable is on the random forest model. Variable importance measures can be considered as machine-learning correlations between input and output parameters.

In this work, we implemented a random forest algorithm from the scikit-learn library [42] for Python. The number of trees was 100; other parameters were left as default, which gave the best performance.

2.3.3. Gradient Boosting

Gradient boosting decision tree methods [43] are also advanced ensemble algorithms that can be based on decision trees (Figure 5). This technique is based on robust agglomeration of additive weak learning models, which iteratively complement each other. The training process of additive models can be represented as follows:

$$F_m(x) = F_{m-1}(x) + \alpha h_m(x) \tag{6}$$

Here, $F_m(x)$ is an agglomeration of models, composed of m weak learning models $h_m(x)$ (i.e., basis functions, in our case, decision trees), that corrects the error of the $F_{m-1}(x)$. To mitigate the model's overfitting, a scaling factor α is applied, which can vary from 0 to 1 in order to decrease the contribution of each iteration. In gradient tree boosting regression, the basic functions $h_m(x)$ are represented by small regression trees. Recently, the gradient tree boosting method has been regularized and implemented in the extreme gradient tree boosting algorithm (XGBoost) [44]. The most significant improvements implemented in XGBoost include loss function regularization and normalization, Taylor expansion enhancement of the loss function, and a more complex split-finding algorithm for many features.

In this study, we used XGBoost Python libraries with 100 iterative, weak learning algorithms; a scaling factor of 0.1; and a learning rate of 0.2. The other parameters were left as default.

2.3.4. Evaluation Criteria

To evaluate the model performance, the root mean square error (*RMSE*) and the mean absolute error (*MAE*) were used:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$$
(7)

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$
(8)

where N is the number of performed tests, and y_i and \hat{y}_i are real and predicted values, respectively.

In addition, the coefficient of determination (also called R^2) was used to evaluate the model performance:

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (y_{i} - \hat{y}_{i})^{2}}{\sum_{i=1}^{n} (y_{i} - \overline{y})^{2}}$$
(9)

where \overline{y} is the mean value of the true data. The coefficient of determination shows how much of the variability in true values can be caused by the relationship to the predicted values. R^2 is represented by a value up to 1: $R^2 = 1$ means the model predicts data with a perfect fit, with 0 the model always predicts the mean value, and negative values mean the model cannot predict the data.

The Pearson method was used for correlation calculations. Pearson's correlation coefficient is defined as:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}$$
(10)

where n is sample size, x_i and y_i are the individual sample points, and \overline{x} and \overline{y} are the sample means. Pearson's correlation coefficient ranges from -1 to 1, where -1 and 1 indicate perfectly negative and positive linear correlations, respectively, and 0 indicates no linear correlation.

To compare Pearson's correlation coefficients with the variable importance measures (which range from 0 to 1; the sum of VIMs gives 1), the fracture toughness correlation

coefficients were normalized by dividing the coefficients by their sum and taking the absolute value:

$$r_{xy_i}^{norm} = \frac{|r_{xy_i}|}{\sum_{i=1}^n r_{xy_i}}$$
(11)

3. Results and Discussions

Overall, to investigate ML prediction capabilities and property correlations, 50 batches of specimens were obtained, 600 specimens were tested, and 900 properties were extracted. The statistics of the mechanical test results are presented in Table 2. The coordinates of the specimen location were also treated as an input parameter, and they were considered as a variation in the environment parameters and as resulting from the human factor during the manufacturing process. Even with this extensive work completed, the dataset would be considered small in the machine-learning field.

Table 2. Statistics and numbering of the analyzed mechanical properties.

#	Property	Min	Max	Mean	STD
0	Coordinates, m	0	50	-	-
1	0° bending failure stress, MPa	239	676	559.2	71.3
2	0° bending modulus, GPa	25.3	34.6	30.3	3.2
3	90° bending failure stress, MPa	14.7	172.0	132.5	25.9
4	90° bending modulus, GPa	5.1	13.8	11.8	1.3
5	0° shear strength, MPa	24.2	58.1	45.6	7.9
6	0° shear modulus, GPa	4.3	9.3	6.9	1.2
7	90° shear strength, MPa	45.2	69.2	54.4	4.9
8	90° shear modulus, GPa	4.9	9.9	7.8	0.7
9	0° compression strength, MPa	366	660	494.4	58.0
10	0° compression modulus, GPa	34.3	54.2	47.4	3.9
11	90° compression strength, MPa	71.6	117.0	91.9	10.1
12	90° compression modulus, GPa	5.2	9.1	6.7	0.8
13	0° impact failure energy, kJ/m ²	30	142	70	27
14	90° impact failure energy, kJ/m ²	9.5	22.5	14.5	3
15	0° tensile strength, MPa	434	839	657.5	85.6
16	0° tensile modulus, GPa	41.9	49.3	45.1	1.5
17	90° tensile strength, MPa	36.4	57.4	45.6	4.5
18	90° tensile modulus, GPa	4.2	6.6	5.4	0.5
19	0° KIC for 40 mm	173.0	305.8	232.4	29.6

The properties were chosen according to their low acquisition time, low costs, and possible correlations with fracture toughness. Fractures in composite materials are usually associated with the matrix properties [45]. We expected a high correlation with bending, impact, and tensile (only in the transverse direction) properties, both for the modulus and strength, because these characteristics are also associated with matrix properties. Other properties were considered for verification of the previous research and hypotheses.

The analysis of mechanical properties indicated anisotropic behavior, which is typical for pultruded materials: the performance at 0° (fiber or longitudinal direction) was much high than at 90° (transverse direction). The material was characterized by relatively high strength (properties 9 and 15) and low Young's modulus, especially in the transverse direction (properties 4, 12, and 18). There were property deviations along the length (due to manufacturing variances), which allowed us to train the ML models on a variable dataset.

Standard deviations of properties were calculated for the obtained dataset. These deviations were mostly observed along the length of the profile, while local deviations were minimal. This was most likely caused by deviations in the conditions for composite production and the human factor. The deviation in fracture toughness was relatively small, considering the measurement difficulties described in Section 2.2.2, and the deviations were lower than those for other fracture toughness measurements due to the experiment's

specifics, as explained in [32]. These deviations were also mostly introduced due to the property differences along the length of the material, and low deviations were observed in one batch (the batch of one meter length). In addition, deviations were considered beneficial for our purposes, since it gave the machine learning more diverse data to learn from.

For additional analysis, the correlation coefficients, representing the relationship between two variables, were obtained using Pearson's method. The heatmap of correlation coefficients for the mechanical properties is shown in Figure 6.



Figure 6. Heatmap of correlations in the acquired dataset. The last row (or column) represents the correlation of the fracture toughness with other properties. Property numeration is the same as in Table 2.

As expected, the fracture toughness stress intensity factor had especially strong correlations with the elastic bending (property 2) and tensile elasticity (property 18) properties, which are associated with matrix properties. Interestingly, the stress intensity factor also had noticeable correlations with the shear strength (properties 7 and 8), the compression modulus in the longitudinal direction (property 10), the compression strength in the transverse direction (property 11), and the tensile strength and modulus in the longitudinal direction (properties 15 and 16). These correlations describe the performance of the material overall and will be discussed with ML variable importance measures. Other properties had low or no correlations with the stress intensity factor.

Three selected machine-learning algorithms were employed: neural networks, a random forest algorithm and XGBoost. Other machine-learning algorithms, such as support vector regression and Gaussian processes (kriging), were not originally designed to handle multidimensional data, and they performed worse in our case. The selected machinelearning models, described in the Materials and Methods section, were trained on the dataset. We used a sensitivity analysis for optimal training procedure where different learning hyperparameters were found for the best prediction results. To test the machine learning methods, we used cross-validation for 45 specimen results for training and 5 specimen results for validation (Figure 7d). The cross-validation was repeated with the entire dataset, and the combined results are presented in Figure 7 and Table 3.



Figure 7. Plots for predicted vs. true values with dots indicating the mean prediction for one data point and vertical lines indicating error bars across different training sessions during cross-validation: (a) neural network (note the different scale for the presentation of error bars), (b) random forest, (c) XGBoost, (d) cross-validation scheme.

Table 3. Prediction results for fracture toughness; the error in relation to the mean experimental results is in brackets.

Model	Mean Prediction RMSE	Mean Prediction MAE	Mean R ²
Neural network	49.1 (21%)	35.0 (15%)	-1.92
Random forest	22.7 (9.8%)	17.0 (7.3%)	0.58
XGBoost	21.8 (9.4%)	16.0 (6.9%)	0.61

As shown in the plots (Figure 7), ensemble-based machine learning methods (random forest, XGBoost) predicted fracture toughness with high accuracy, with 9.8% and 9.4% errors in relation to the mean experimental results, which were within the experimental error of 12.7%. The accuracy was also confirmed by an R^2 coefficient above 0.5. Despite thorough selection of the architecture, neural networks showed the worst performance due to the small dataset used; it was challenging to optimize all the NN parameters when we did not have enough data. Random forest and XGBoost showed similar promising results. XGBoost slightly outperformed random forest in the RMSE, MAE, and R^2 metrics.

Random forest and XGBoost allow the calculation of variable importance measures (Figure 8), which show how the input feature influence the results during the optimization process.



Figure 8. Variable importance vs. normalized correlations: (a) random forest, (b) XGBoost, (c) normalized correlations. Properties are assigned similarly as in Table 2.

Variable importance measure analysis showed that random forest and XGBoost used different variables in their predictions: random forest mostly leaned on the bending modulus in the 0° direction (property 2), while XGBoost leaned on the tensile modulus in the 90° direction (property 18).

It should be noted that both algorithms, random forest and XGBoost, considered the longitudinal compression modulus and tensile strength (properties 10 and 15) as the second most important variables to predict fracture toughness, despite their having the same correlations as many other characteristics (for example, properties 7 and 11). The influence of the compression modulus and tensile strength on fracture properties is not straightforward or intuitive. However, the influence of compression and tension can be explained by local strain/stress fields around fibers during crack development. Our theory is that, during the test, before energy is released and the crack grows, energy accumulates at locations under compression. This hypothesis can be confirmed by consulting previous studies; for instance, those by Tsouvalis et al. [46] or Song et al. [47], where the authors discuss the fiber–matrix interface and observe compression strain and stress fields during fracture simulation. This led us to the assumption that machine learning can find mechanical correlations at the micro-level, despite only knowing the macro-characteristics. However this theory needs to be researched further and the machine-learning transition from macro- to micro-parameters needs to be specifically investigated.

To further investigate the ML prediction performance and the correlations of the fracture toughness and standard properties, we selected the five most significant properties from the analysis of both variable importance measures; for the random forest: 0, 2, 5, 10, and 15; for XGBoost: 2, 7, 10, 15, and 18. Considering that some properties were repeated, overall, we selected 0, 2, 5, 7, 10, 15, and 18, which had high or moderate correlations with the fracture toughness stress intensity factor.

With these seven properties, the machine-learning algorithms were trained again, and the results are presented in Figure 9 and Table 4.



Figure 9. Predicted vs. true plots for selected features, with dots indicating the mean predictions for individual data points and vertical lines indicating error bars across different training sessions during cross-validation: (a) Neural network (note the different scale for the presentation of the error bars), (b) random forest, (c) XGBoost.

Table 4. Prediction results for selected features of fracture toughness; the error in relation to the mean experimental results is in brackets.

Model	Mean Prediction RMSE	Mean Prediction MAE	Mean R ²	
Neural network	28.4 (12.2%)	24.0 (10.3%)	-1.15	
Random forest	23.4 (10.1%)	17.5 (7.5%)	0.52	
XGBoost	23.8 (10.2%)	18.8 (8.1%)	0.54	

With the selected properties and optimized architecture, the neural network's performance strongly increased. Therefore, we could exploit a simpler architecture for fewer features, which would be trained better on small datasets. However, the prediction accuracies of random forest and XGBoost fell slightly, especially the accuracy of the results for XGBoost because, according to its VIMs, the algorithm used almost all the properties for precise predictions and dropped properties could not be ignored. It is worth noting that, with the selected features, all the ML algorithms predicted the fracture toughness within the experimental error range.

In the future, the presented approach will be extended for predictions and correlation analysis of different mechanical properties and different composite materials. A larger dataset will be acquired from different manufacturers, other production methods, and composite components. The correlations revealed here will be verified, and a pre-trained model for the general case of composite materials will be developed. Furthermore, the

13 of 15

present methodology will be applied to cut back the mechanical testing study required for the comprehensive characterization of composite materials.

4. Conclusions

This paper proposes a data-driven approach for prediction of the fracture toughness mechanical property based on other properties of the material. Furthermore, correlations of easy-to-measure mechanical properties with fracture toughness were investigated based on machine-learning algorithms. To illustrate the proposed approach, three machine-learning models were implemented and trained: an artificial neural network, a random forest algorithm, and gradient boosting decision trees. A considerable dataset (900 properties, 50 batches) of mechanical properties for a pultruded composite material was obtained to train the models.

Machine learning proved its ability to predict fracture toughness behavior in the absence of information about the inner microstructure. The analysis of the ML models showed that the gradient boosting model predicted the stress intensity factor with an MSE less than 10%, which was equivalent to the experimental error. The random forest algorithm showed similar performance. Prediction of fracture toughness with the neural network method was considered statistically unsatisfactory due to the small database used for the significant number of trainable parameters that had to be optimized during training.

Feature selection and correlation analysis showed that some properties correlated with the material's fracture toughness more than others: elastic bending and tensile elasticity correlated well with fracture toughness, as expected, due to the nature of the matrix behavior, and a good correlation with fracture toughness was also observed for the longitudinal compression modulus and tensile strength, which could have been caused by energy accumulations at locations under compression. However, future investigation is required. Machine learning only with selected features with high correlations significantly improved the neural network predictions but lowered the accuracy of the ensemble-based algorithms.

Overall, machine-learning algorithms show potential for determining which mechanical characteristics at the micro-level are correlated with the macro-parameters without knowing the internal microstructure. The data-driven approach for mechanical property prediction can complement and enhance physics-based prediction methods and lead to a cut back in the experimental framework required to characterize composite material extensively for structural applications.

Author Contributions: Conceptualization: R.K., I.A. and I.V.S.; Methodology: R.K., I.A. and I.V.S.; Software: R.K.; Validation: R.K.; Formal analysis: R.K. and I.V.S.; Investigation: R.K. and I.V.S.; Resources: I.V.S.; Data curation: I.V.S.; Writing—original draft: R.K.; Writing—review and editing: R.K., I.A. and I.V.S.; Visualization: R.K.; Supervision: I.A. and I.V.S.; Project administration: I.A.; Funding acquisition: I.A. and I.V.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded in the framework of the "Experimental and digital certification platform" project (#400-248) of the National Technological Initiative.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data used in this research will be made available on request.

Acknowledgments: The authors thank S. Gusev, under whose supervision the pultruded composite material was manufactured at the Laboratory of Composite Materials and Structures, CMT, Skoltech. The authors also thank Stepan Konev, Ruslan Vorobyev, and Boris Voloskov, engineers at the Mechanical Testing Laboratory, CMT, Skoltech.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Wang, J.; GangaRao, H.; Liang, R.; Liu, W. Durability and prediction models of fiber-reinforced polymer composites under various environmental conditions: A critical review. *J. Reinf. Plast. Compos.* **2015**, *35*, 179–211. [CrossRef]
- Ma, Z.; Zhang, P.; Zhu, J. Review on the fatigue properties of 3D woven fiber/epoxy composites: Testing and modelling strategies. J. Ind. Text. 2020, 51 (Suppl. S5), 7755S–7795S. [CrossRef]
- Baran, I.; Cinar, K.; Ersoy, N.; Akkerman, R.; Hattel, J.H. A Review on the Mechanical Modeling of Composite Manufacturing Processes. Arch. Comput. Methods Eng. 2016, 24, 365–395. [CrossRef] [PubMed]
- 4. Bostanabad, R.; Zhang, Y.; Li, X.; Kearney, T.; Brinson, L.C.; Apley, D.W.; Liu, W.K.; Chen, W. Computational microstructure characterization and reconstruction: Review of the state-of-the-art techniques. *Prog. Mater. Sci.* **2018**, *95*, 1–41. [CrossRef]
- 5. Mohan, N.; Senthil, P.; Vinodh, S.; Jayanth, N. A review on composite materials and process parameters optimisation for the fused deposition modelling process. *Virtual Phys. Prototyp.* **2017**, *12*, 47–59. [CrossRef]
- Zhou, X.-Y.; Gosling, P.; Pearce, C.; Kaczmarczyk, L.; Ullah, Z. Perturbation-based stochastic multi-scale computational homogenization method for the determination of the effective properties of composite materials with random properties. *Comput. Methods Appl. Mech. Eng.* 2015, 300, 84–105. [CrossRef]
- Karabasov, S.; Nerukh, D.; Hoekstra, A.; Chopard, B.; Coveney, P.V. Multiscale modelling: Approaches and challenges. *Philos. Trans. R. Soc. London. Ser. A Math. Phys. Eng. Sci.* 2014, 372, 20130390. [CrossRef] [PubMed]
- Hoekstra, A.; Chopard, B.; Coveney, P. Multiscale modelling and simulation: A position paper. *Philos. Trans. R. Soc. London. Ser. A Math. Phys. Eng. Sci.* 2014, 372, 20130377. [CrossRef]
- 9. Chen, C.-T.; Gu, G.X. Machine learning for composite materials. MRS Commun. 2019, 9, 556–566. [CrossRef]
- 10. Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine learning in materials informatics: Recent applications and prospects. *NPJ Comput. Mater.* **2017**, *3*, 1. [CrossRef]
- 11. Mukherjee, A.; Schmauder, S.; Rugle, M. Artificial neural networks for the prediction of mechanical begavior of metal matrix composites. *Acta Met.* **1995**, *43*, 4083–4091. [CrossRef]
- 12. Koker, R.; Altinkok, N.; Demir, A. Neural network based prediction of mechanical properties of particulate reinforced metal matrix composites using various training algorithms. *Mater. Des.* 2007, 28, 616–627. [CrossRef]
- 13. Altinkok, N.; Koker, R. Neural network approach to prediction of bending strength and hardening behaviour of particulate reinforced (Al–Si–Mg)-aluminium matrix composites. *Mater. Des.* **2004**, *25*, 595–602. [CrossRef]
- 14. Sha, W.; Edwards, K. The use of artificial neural networks in materials science based research. *Mater. Des.* **2007**, *28*, 1747–1752. [CrossRef]
- 15. Yu, W.; Li, M.; Luo, J.; Su, S.; Li, C. Prediction of the mechanical properties of the post-forged Ti–6Al–4V alloy using fuzzy neural network. *Mater. Des.* **2010**, *31*, 3282–3288. [CrossRef]
- Ye, S.; Li, B.; Li, Q.; Zhao, H.-P.; Feng, X.-Q. Deep neural network method for predicting the mechanical properties of composites. *Appl. Phys. Lett.* 2019, 115, 161901. [CrossRef]
- 17. Yang, C.; Kim, Y.; Ryu, S.; Gu, G.X. Prediction of composite microstructure stress-strain curves using convolutional neural networks. *Mater. Des.* **2020**, *189*, 108509. [CrossRef]
- Qi, Z.; Zhang, N.; Liu, Y.; Chen, W. Prediction of mechanical properties of carbon fiber based on cross-scale FEM and machine learning. *Compos. Struct.* 2019, 212, 199–206. [CrossRef]
- 19. Jiang, Z.; Zhang, Z.; Friedrich, K. Prediction on wear properties of polymer composites with artificial neural networks. *Compos. Sci. Technol.* 2007, 67, 168–176. [CrossRef]
- 20. Jiang, Z.; Gyurova, L.; Zhang, Z.; Friedrich, K.; Schlarb, A.K. Neural network based prediction on mechanical and wear properties of short fibers reinforced polyamide composites. *Mater. Des.* **2008**, *29*, 628–637. [CrossRef]
- Shirvanimoghaddam, K.; Khayyam, H.; Abdizadeh, H.; Akbari, M.K.; Pakseresht, A.; Ghasali, E.; Naebe, M. Boron carbide reinforced aluminium matrix composite: Physical, mechanical characterization and mathematical modelling. *Mater. Sci. Eng. A* 2016, 658, 135–149. [CrossRef]
- 22. Li, X.; Liu, Z.; Cui, S.; Luo, C.; Li, C.; Zhuang, Z. Predicting the effective mechanical property of heterogeneous materials by image based modeling and deep learning. *Comput. Methods Appl. Mech. Eng.* **2019**, 347, 735–753. [CrossRef]
- Gu, G.X.; Chen, C.-T.; Buehler, M.J. De novo composite design based on machine learning algorithm. *Extrem. Mech. Lett.* 2018, 18, 19–28. [CrossRef]
- 24. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), Savannah, GA, USA, 2–4 November 2016.
- Yang, Z.; Yabansu, Y.C.; Jha, D.; Liao, W.-K.; Choudhary, A.N.; Kalidindi, S.R.; Agrawal, A. Establishing structure-property localization linkages for elastic deformation of three-dimensional high contrast composites using deep learning approaches. *Acta Mater.* 2018, *166*, 335–345. [CrossRef]
- 26. Tiryaki, S.; Aydın, A. An artificial neural network model for predicting compression strength of heat treated woods and comparison with a multiple linear regression model. *Constr. Build. Mater.* **2014**, *62*, 102–108. [CrossRef]
- 27. Lu, L.; Dao, M.; Kumar, P.; Ramamurty, U.; Karniadakis, G.E.; Suresh, S. Extraction of mechanical properties of materials through deep learning from instrumented indentation. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 7052–7062. [CrossRef]
- 28. Mottram, J.T. Compression Strength of Pultruded Flat Sheet Material. J. Mater. Civ. Eng. 1994, 6, 185–200. [CrossRef]

- Alqam, M.; Bennett, R.M.; Zureick, A.-H. Three-parameter vs. two-parameter Weibull distribution for pultruded composite material properties. *Compos. Struct.* 2002, 58, 497–503. [CrossRef]
- Vedernikov, A.; Safonov, A.; Tucci, F.; Carlone, P.; Akhatov, I. Pultruded materials and structures: A review. J. Compos. Mater. 2020, 54, 4081–4117. [CrossRef]
- 31. El-Hajjar, R.; Haj-Ali, R. Mode-I fracture toughness testing of thick section FRP composites using the ESE(T) specimen. *Eng. Fract. Mech.* **2005**, *72*, 631–643. [CrossRef]
- 32. Fernandes, L.A.; Silvestre, N.; Correia, J.R. Characterization of transverse fracture properties of pultruded GFRP material in tension. *Compos. Part B Eng.* 2019, 175, 107095. [CrossRef]
- Almeida-Fernandes, L.; Silvestre, N.; Correia, J.R.; Arruda, M. Fracture toughness-based models for damage simulation of pultruded GFRP materials. *Compos. Part B Eng.* 2020, 186, 107818. [CrossRef]
- 34. *ASTM International E399-12;* Standard Test Method for Linear-Elastic Plane-Strain Fracture Toughness K Ic of Metallic Materials. ASTM: West Conshohocken, PA, USA, 2012; pp. 1–33. [CrossRef]
- Nwankpa, C.; Ijomah, W.; Gachagan, A.; Marshall, S. Activation Functions: Comparison of trends in Practice and Research for Deep Learning. *arXiv* 2018, arXiv:1811.03378.
- 36. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
- Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
- Segal, M.R. Machine Learning Benchmarks and Random Forest Regression, Biostatistics; UCSF: San Francisco, CA, USA, 2004; pp. 1–14. Available online: http://escholarship.org/uc/item/35x3v9t4.pdf (accessed on 24 January 2022).
- 39. Sexton, J.; Laake, P. Standard errors for bagged and random forest estimators. *Comput. Stat. Data Anal.* 2009, 53, 801–811. [CrossRef]
- 40. Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* 2007, *8*, 25. [CrossRef]
- 41. Archer, K.J.; Kimes, R.V. Empirical characterization of random forest variable importance measures. *Comput. Stat. Data Anal.* **2008**, *52*, 2249–2260. [CrossRef]
- 42. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- 43. Friedman, J.H. Stochastic gradient boosting. Comput. Stat. Data Anal. 2002, 38, 367–378. [CrossRef]
- Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. [CrossRef]
- 45. Miyajima, T.; Sakai, M. The fracture toughness for first matrix cracking of a unidirectionally reinforced carbon/carbon composite material. *J. Mater. Res.* **1991**, *6*, 2312–2317. [CrossRef]
- Tsouvalis, N.G.; Anyfantis, K.N. Determination of the fracture process zone under Mode I fracture in glass fiber composites. J. Compos. Mater. 2011, 46, 27–41. [CrossRef]
- 47. Song, L.; Meng, S.; Xu, C.; Fang, G.; Yang, Q. Finite element-based phase-field simulation of interfacial damage in unidirectional composite under transverse tension. *Model. Simul. Mater. Sci. Eng.* **2019**, 27, 55011. [CrossRef]