Machine Learning with Applications 19 (2025) 100609

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

# Application of convolutional neural networks and ensemble methods in the fiber volume content analysis of natural fiber composites

Florian Rothenhäusler [a,1], Rodrigo Queiroz Albuquerque [a,1], Marcel Sticher [a],
Christopher Kuenneth [b], Holger Ruckdaeschel [a,*]

[a] Department of Polymer Engineering, University of Bayreuth, Universitätsstrasse 30, Bayreuth 95447, Bavaria, Germany
[b] Faculty of Engineering Science, University of Bayreuth, Universitätsstrasse 30, Bayreuth 95447, Bavaria, Germany

## ARTICLE INFO

Dataset link: https://github.com/Polymer-Engineering-University-Bayreuth/FiberVolumeContent_ConvNet

*Keywords:*
Flax fibers
Sustainability
Machine learning
Image detection
Natural fiber composites
CNN

## ABSTRACT

The incorporation of natural fibers into fiber-reinforced polymer composites (FRPC) has the potential to bolster their sustainability. A critical attribute of FRPC is the fiber volume content (*FVC*), a parameter that profoundly influences their thermo-mechanical characteristics. However, the determination of *FVC* in natural fiber composites (NFC) through manual analysis of light microscopy images is a labor-intensive process. In this work, it is demonstrated that the pixels from light microscopy images of NFC can be utilized to predict *FVC* using machine learning (ML) models. In this proof-of-concept investigation, it is shown that convolutional neural network-based models predict *FVC* with an accuracy required in polymer engineering applications, with a mean average error of 2.72 % and an $R^2$ coefficient of 0.85. Finally, it is shown that much simpler ML models, non-specialized in image recognition, besides being much easier and more efficient to optimize and train, can also deliver good accuracies required for *FVC* characterization, which not only contributes to the sustainability, but also facilitates the access of such models by researchers in regions with little computational resources. This study marks a substantial advancement in the area of automated characterization of NFC, and democratization of knowledge, offering a promising avenue for the enhancement of sustainable materials.

## 1. Introduction

Fiber-reinforced polymer composites (FRPC) find extensive use in automotive, wind energy generation, and aerospace applications, primarily due to their high weight-specific Young's modulus and strength (Ashby, 1992; Ehrenstein, 2006), which are attributed to the reinforcing fibers (Flemming et al., 1995; Schürmann, 2007). Traditionally, synthetic fibers such as glass and carbon have been utilized as reinforcements for FRPC. The production of these fibers involves substantial energy consumption, leading to relevant $CO_2$ emissions (Das, 2011). In contrast, plant-based fibers require less energy for production and processing (Bos, 2004). Additionally, natural fibers are biodegradable, presenting an environmentally friendly alternative to glass and carbon fibers. Plant-based fibers, like flax and hemp, also have lower density and cost compared to synthetic fibers (Bledzki & Gassan, 1997). However, plant-based fibers exhibit lower strength and modulus than synthetic fibers (Flemming et al., 1995). Despite this, they can be a sustainable alternative in applications where the highest strength and modulus are not essential.

The mechanical properties of FRPC, influenced by factors such as fiber length (Fu & Lauke, 1996; Gupta et al., 1989), fiber orientation (Chin et al., 1988; Fu & Lauke, 1998), and fiber volume content (*FVC*) (Affdl & Kardos, 1976), are crucial in diverse engineering contexts. The *FVC*, denoting the ratio of fiber volume to the total volume in a FRPC (Schürmann, 2007), is particularly integral in determining specific mechanical properties (Voigt, 1889) and is expressed by Eq. (1):

$$FVC = \frac{V_{\text{fiber}}}{V_{\text{composite}}} \tag{1}$$

For composites reinforced with glass or carbon fibers, *FVC* analysis is typically conducted through techniques like thermo-gravimetric analysis (TGA) (DIN EN ISO 1172:1998-12, 1998; Grund et al., 2019), density measurements (Abdalla et al., 2008), computed tomography (CT) (Hessman et al., 2019), or light microscopy (LM) (Jock, 1986). However, natural fiber composites (NFC) pose challenges due to their organic nature, low thermal stability, and characteristic hollow structure (Kumar et al., 2022; Mwaikambo & Ansell, 2001). Analysis

---

[*] Corresponding author.
*E-mail addresses:* florian.rothenhaeusler@uni-bayreuth.de (F. Rothenhäusler), rodrigo.q.albuquerque@uni-bayreuth.de (R.Q. Albuquerque), marcel.sticher@uni-bayreuth.de (M. Sticher), christopher.kuenneth@uni-bayreuth.de (C. Kuenneth), holger.ruckdaeschel@uni-bayreuth.de (H. Ruckdaeschel).
[1] Contributed equally to this work.

https://doi.org/10.1016/j.mlwa.2024.100609
Received 23 September 2024; Received in revised form 15 November 2024; Accepted 26 November 2024

methods such as TGA may be hindered by the low thermal stability of natural fibers, especially when additives with low thermal stability are present in the matrix. Similarly, the hollow structure of natural fibers complicates *FVC* determination through density measurements, and the organic nature makes precise CT detection challenging. Consequently, LM analysis of polished cross-sections of unidirectional NFC is often preferred.

Manual analysis of LM images via image processing software proves to be laborious and time-intensive. In contrast, artificial neural networks (ANN), such as convolutional neural networks (CNN), offer a faster and precise alternative once properly trained (Andrade & Eduardo, 2011; Chauhan et al., 2018). Previous applications of neural networks have demonstrated substantial reductions in workload, exemplified in defect detection in metal screws and the inline assessment of part porosity in thermoplastic polymers produced via powder-based additive manufacturing (Schlicht et al., 2022; Song et al., 2018).

Different machine learning (ML) models have been used to model mechanical properties of polymer composites, as highlighted in the recent review by Sorour et al. (2024). Machello et al. (2024) trained decision tree-based models using features like the exposure temperature, $T_g$ and sample thickness to predict the tensile strength retention of a FRPC material, achieving $R^2$ of 0.88–0.96 for the test set. Osa-Uwagboe et al. (2024) screened several ML models that used quasi-static indentation data as features to predict damage properties in FRPC materials, achieving $R^2$ larger than 0.98 for the test set. Shokrollahi et al. (2023) have employed deep CNNs to predict local stress fields within 2D slices of composite microstructure images, revealing that this ML-based technique can be an alternative to costly force electron microscopy measurements.

The importance of determining *FVC* via ML models and image recognition lies in their much faster, cheaper and more sustainable nature as compared to traditional destructive methods. To the best of our knowledge, no paper has been found describing the use of ML models to determine *FVC* of natural fiber composites from LM images. However, one paper has been found describing the use of CNN models to determine *FVC* of carbon fiber composites (Blarr et al., 2024) from 3D CT images. This model exhibited high variance and was trained using a very small experimental data set (14 samples) with a small range of *FVC* (22%–29%), and the trained CNN showed a very low $R^2$ score of −0.11, which is apparently related to the training error. The current investigation reveals that it is possible to train an accurate CNN model with a *FVC* range five times larger and still get high $R^2$ coefficients and small errors for the prediction of this property on both the training and test sets.

This proof-of-concept study aims at demonstrating that pixels from LM images of NFCs can be used to train CNN models to predict *FVC* with accuracies good enough for being used in usual polymer engineering applications. It is also shown that comparatively simpler ML models like the eXtreme Gradient Boosting (XGBoost), which are much easier and faster to optimize and train, but which are not specialized in image recognition problems, can also predict *FVC* and deliver an accuracy equally acceptable for *FVC* characterizations found in polymer engineering problems. Both models are trained using LM images of an NFC produced through the hand-layup process involving flax fibers and an epoxy resin matrix. Afterwards, the performance of the ML models is improved by employing ensemble techniques.

## 2. Methods

### 2.1. Materials

The diglycidyl ether of bisphenol F (DGEBF) utilized in this study, exhibiting an epoxide equivalent weight (EEW) of $169\,\mathrm{g\,mol^{-1}}$, was obtained in the form of Epilox F17-00 from LEUNA-Harze GmbH, located in Leuna, Germany. The diglycidyl ether of bisphenol A (EEW = $187\,\mathrm{g\,mol^{-1}}$) was sourced from Blue Cube Assets GmbH & Co. KG, Olin Epoxy, based in Stade, Germany. The Heloxy modifier BD was procured from Westlake Epoxy GmbH, headquartered in Duisburg, Germany. Diethyl toluene diamine (DETDA), characterized by an active hydrogen equivalent weight of $44.57\,\mathrm{g\,mol^{-1}}$, was provided by Schill+Seilacher GmbH, situated in Böblingen, Germany, and was supplied in the form of Polyvertec®H81. Jeffamine®D-230 was obtained from Huntsman Corporation, located in Salt Lake City, US.

For the natural reinforcement component, non-crimp flax fabric FUSE 200 FLX was employed, sourced from SachsenLeinen GmbH in Markkleeberg, Germany. FUSE 200 FLX comprises unidirectional, non-twisted yarns bound together by a thermoplastic fiber binder, constituting approximately 5% of the total fabric weight.

### 2.2. Resin formulation & laminate curing

DGEBF and DETDA were weighed in so that the stoichiometric ratio between the active hydrogen atoms in DETDA to the epoxy groups in DGEBF is equal to one. The blending of DGEBF and DETDA involved the utilization of a centrifuge speed mixer from Hauschild Engineering, located in Hamm, Germany, operating at $3000\,\mathrm{min^{-1}}$ for a duration of 120 s.

Flax fibers, cut to lengths of 10 cm, were manually impregnated with the resin. After aligning the fibers uni-directionally and placing them within a vacuum bag (maintained at approximately 1 mbar), the composite underwent curing in a Memmert ULE 400 convection oven from Memmert GmbH + Co. KG, based in Schwabach, Germany. The curing process consisted of one hour at 120 °C followed by three hours at 160 °C. Subsequently, the composite was gradually cooled to room temperature over a 4-h duration.

The resulting composite was then sectioned into pieces measuring 15 mm by 15 mm using a Mutronic DIADISC5200 diamond plate saw from MUTRONIC Präzisionsgerätebau GmbH & Co. KG, located in Rieden am Forggensee, Germany. These specimens were embedded in a matrix composed of DGEBA, Heloxy modifier BD, and Jeffamine®D-230, maintaining a stoichiometric ratio between the functional groups. The embedding matrix underwent a curing process lasting 12 h at 40 °C in a Memmert ULE 400 convection oven from Memmert GmbH + Co. KG.

### 2.3. Grinding & polishing

The preparation of specimens involved grinding and polishing procedures using a Struers RotoPol-21 equipped with a Struers RotoForce-4, both provided by Struers GmbH in Ottensoos, Germany. For the grinding process, silicon carbide papers from Struers with varying grain sizes ranging from 68 μm to 10 μm were employed. Each grinding step, lasting 10 min, applied a force of 10 N to each specimen. The specimens were rotated in the opposite direction to the silicon carbide papers, which rotated at 300 rpm. Subsequently, polishing was carried out using a diamond dispersion featuring 6 μm grain sizes, with the specimens rotating at 150 rpm under a force of 5 N. The final step involved cleaning the specimens in an ultra-sonic bath utilizing de-ionized water.

### 2.4. Light microscopy

The analysis of the composite's cross-section was conducted using a Leica DM 6000M microscope from Leica Microsystems GmbH in Wetzlar, Germany, equipped with a lens providing 50 times magnification. The resulting images had dimensions of 200 μm by 150 μm (width × height). Gray scale images were captured in bright field mode, utilizing a brightness setting of 85%, exposure time of 16.8 ms and a gamma value of 1.3. The images were randomly acquired throughout the entire cross-section of the NFC and saved as TIFF files. Exemplary images are presented in Fig. 3(a). The distribution of *FVC* across all 290 recorded images is illustrated in Fig. 3(b).

### 2.5. Image analysis

The analysis of LM images was conducted using a routine programmed in ImageJ image processing software (ImageJ 2.14.0/1.54f). However, automated processing was hampered by inhomogeneous illumination of the images due to the light microscopes' illumination setup. The light cone renders the center of the image darker than the border areas resulting in problems in setting the threshold between gray scale values of matrix and fibers. Therefore, an artificial background was created beforehand by overlaying several areas of images where no fibers are present to obtain an image with only matrix. On this image a Gaussian blur filter was applied using ImageJ with a sigma (radius) of 100 to average over local gray value fluctuations. The image processing routine performs the following steps:

1. Directory: creating a directory where the raw images are stored, one where the processed images are going to be stored and inquiring an artificial background. The filenames are stored in a list and a for loop iterates in the range of the length of the list, i.e. over all files.
2. Background correction: subtracting background from raw images and saves the result in 5_ramps LUT (pseudo-color image look-up tables) (ImageJ, 2023) to check the illumination homogeneity.
3. Enhance contrast: increasing contrast by normalization and equalization of pixel gray value histogram with saturated pixels set to 0.1%.
4. Thresholding: setting a threshold by localizing the local minimum between the fiber and the matrix peak in the gray scale histogram using the 'minimum' algorithm. Results in a binary image with white features and black background (matrix), which is saved.
5. Denoising: The despeckle tool is applied iteratively to eliminate artifacts.
6. Eroding and Dilating: Erodes (removes pixels) and dilates (adds pixels) all features (white) in the binary image using the same number of iterations for both operations. This removes features that measure only few pixels in at least one dimension, e.g. grind marks and dust, while retaining the shape and size of larger structures, i.e. the fibers.
7. Measure: The area fraction (% of white pixels) of the binary image is measured and appends it together with the filename to a list.
8. Overlay: The original image and the binary image are overlaid and flattened. The resulting image is converted to 8-bit format and saved. The image processing efficacy was assessed by visually evaluating the mapping accuracy of the overlay images.

This standardized procedure was consistently applied to all images analyzed (see Fig. 1). In this context, *FVC* was quantified through image analysis, where Eq. (2) expresses the ratio of pixels corresponding to fibers (white) to the total number of pixels in the image:

$$FVC = \frac{n_{\text{fiber}}}{n_{\text{total}}} \tag{2}$$

The unmodified 16-bit .tiff images, along with the calculated *FVC*, were provided as input to the CNN.

### 2.6. Convolutional neural network

*Background.* CNNs are an special type of neural networks specialized in image recognition. It basically works as follows. A small patch, typically of size 3 by 3 neurons, slides over the original image, operating on the current nine pixel intensities ($I_i$) that fall inside this patch, according to

$$\sigma(w_1 I_1 + w_2 I_2 + \cdots + w_9 I_9 + b) \tag{3}$$

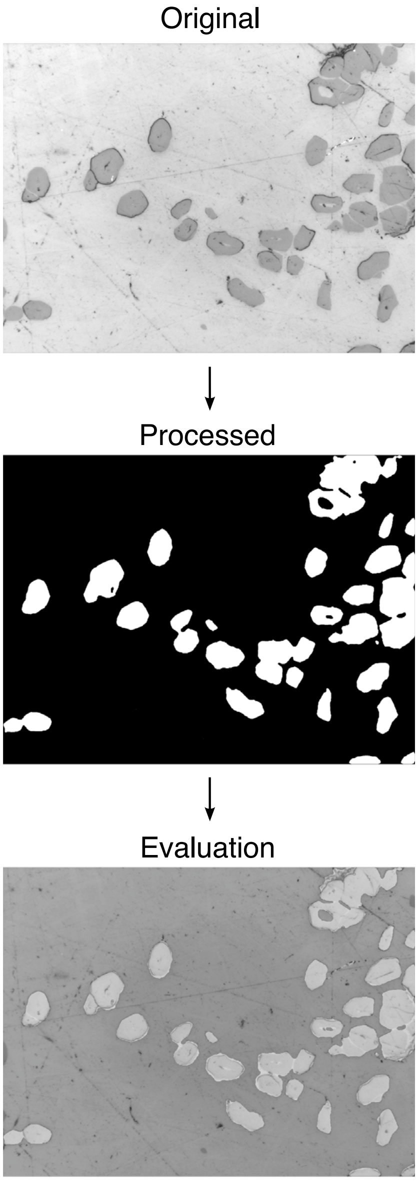

**Fig. 1.** Exemplary image during various stages of the analysis.

where $w_i$'s are weight coefficients which are kept constant each time the patch slides over the image, and $b$ is an added bias (a constant). The result of this operation is sent through an activation function (here, the sigmoid function, $\sigma$), so that a final number (or output) is generated. In Eq. (3), $\sigma(x)$ is given by

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{4}$$

In Eq. (4), a number between zero and one is generated, which is associated with the probability of detecting a specific pattern. Each time the patch slides to a different location, it processes nine different neurons and a different output is generated, which is stored in the so-called feature map. Each convolutional layer can have many feature maps corresponding to the detection of different (local or global) patterns. Local patterns are detected in the first convolutional layers, while global patterns are detected after more convolutional layers have been applied. The last convolutional layer is in general flattened, i.e., all neurons are depicted as a very long 1D vector instead of many (2D) planes of neurons. Following some fully-connected layers of neurons, each with different weight coefficients and activation functions, a final

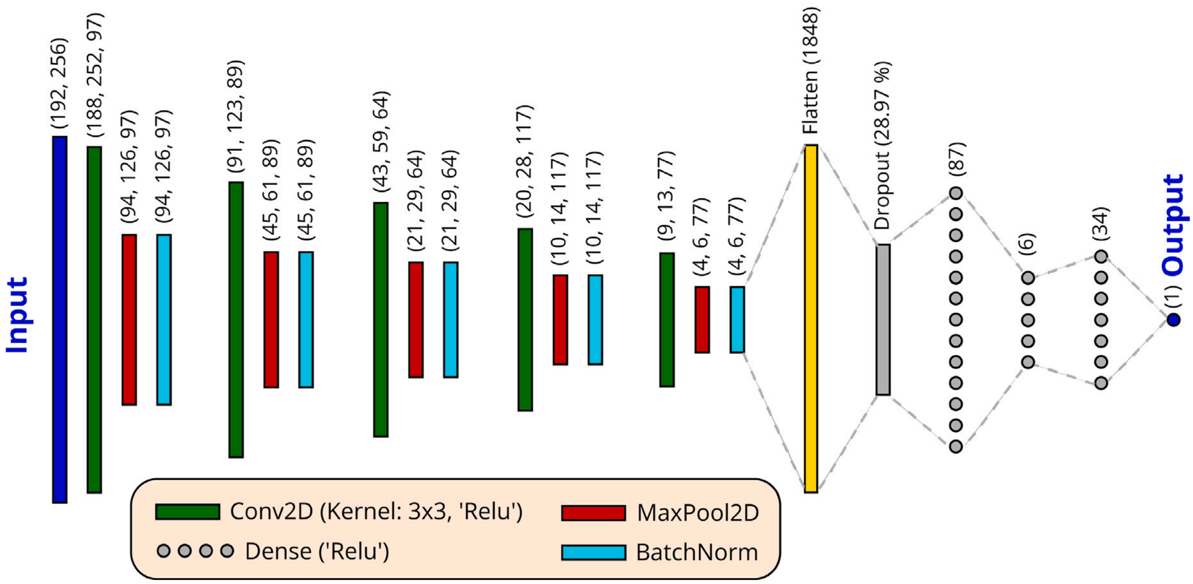

**Fig. 2.** Optimized architecture of the CNN model. The number of neurons in each layer is shown in parenthesis (except for the dropout layer), where (a,b,c) = (height, width, number of filters). A batch of 24 samples and the *MSE* loss function were used. Total number of trainable parameters = 432,301.

layer called output layer is used to predict the final property (*FVC*). Because each feature map has been generated using the same few weight coefficients, CNNs are not fully-connected neural networks and have therefore much less parameters than fully-connected NNs, which renders the former more computationally efficient towards complex tasks involving image recognition.

*Dataset and preprocessing*. All microscopy images were used in gray scale and had resolution of 256 × 192 pixels (width × height), besides being normalized to have their pixel intensities in the range of [0,1]. The dataset, composed of 290 images, was split into training set (80% of data or 232 images) and test set (58 images) using the train_test_split function of the sklearn package and the default random seed number (of 42) shown in the documentation. The training set was augmented to 2552 images (10 augmented images per original image), where each original image was either rotated (5–15 degrees), horizontally or vertically flipped, or horizontally or vertically translated (up to 60% of the total width). During the rotations, the pixels leaving the area of the image were re-inserted to preserve *FVC*. The features used to train the ML models were the pixel intensities, while the manually determined *FVC* (see Section 2.5) was used as the target property. Standardization was used to preprocess *FVC*.

*Metrics*. The CNN model was evaluated via the mean absolute error (*MAE*, see Eq. (5)), mean absolute percentage error (*MAPE*, see Eq. (6)), mean squared error (*MSE*, see Eq. (7)), and the $R^2$ score (see Eq. (8)), which are given by:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \mid y_i - \hat{y}_i \mid \tag{5}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \tag{6}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{7}$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{8}$$

where $n$ is the number of samples, $y_i$ and $\hat{y}_i$ are the true and predicted target property for sample $i$, respectively, and $\bar{y}$ is the average target property.

*Model optimization*. The CNN model was created using tensorflow (Abadi et al., 2015) and Python 3. The hyperparameters were optimized using the optuna library (Akiba et al., 2019), which minimized the average *MAE* evaluated via 5-fold cross validation using only the training set. The following hyperparameters were optimized: total number of convolutional layers (1–5), number of feature maps in each convolutional layer (2–128), activation function (“sigmoid”, “relu”, “selu” and “tanh”), batch size (4–128), size of the dropout layer (0–50%), patch kernel sizes (3 × 3 or 5 × 5) and the loss function used to train the model (“mae” or “mse”). The ADAM optimizer was adopted here. Early stopping and a dropout layer were used in the model to try to minimize overfitting. The final optimized architecture of the CNN model is shown in Fig. 2. After the hyperparameter optimization, a fresh CNN model was trained using the whole training set and predictions were performed on the test set.

*Comparison with other models*. There are numerous state-of-the-art CNN-based models, such as those from the ResNet (He et al., 2016), VGGNet (Simonyan & Zisserman, 2015), and Inception Networks (Szegedy et al., 2016) families, that have the potential to outperform the vanilla CNN model explored in this study. However, the primary aim of this proof-of-concept investigation is not to exhaustively maximize accuracy by employing these complex models through transfer learning. Given that the accuracy achieved — especially in the ensemble model presented in the results — already meets the requirements for polymer engineering, we opted to compare the vanilla CNN model with simpler ML models that are not specialized in image recognition tasks. These simpler models are significantly faster, easier to optimize, and much more efficient to train, making them a viable alternative for predicting *FVC* in regions with limited computational resources. Additionally, the efficiency in training these simpler models contributes to higher sustainability. The XGBoost (eXtreme Gradient Boosting Chen & Guestrin, 2016) model was then chosen for this model comparison, as it was the only simple ML model showing a good prediction on the test set. XGBoost is a highly popular and powerful ML algorithm utilized for supervised learning tasks, particularly in predictive modeling and classification problems. It belongs to the ensemble learning family, which involves combining the predictions of multiple individual models (here, decision trees), to create a stronger and more accurate final model. XGBoost employs gradient boosting to construct a sequence of decision trees, where each tree corrects the errors made by its predecessor using gradient descent optimization. This algorithm has L2 (or Ridge) regularization by default for help minimizing overfitting, and parallel and distributed computing for speed and efficiency. In the present work, the number of estimators (or trees), the learning rate, the maximum depth of each tree, and the L1 and L2 regularization parameters were optimized via a grid search using the training set and a 5-fold cross-validation approach. As XGBoost is not specialized in image recognition, the pixels of the 2D images were first flattened to a long 1D vector of 49 152 components and every fifth component (or pixel) of this vector was used for training and testing the model, as this is known to increase the model efficiency (Albuquerque et al., 2024). After the hyperparameter optimization, a fresh XGBoost model was trained using the whole training set and predictions were performed on the test set.

*Ensemble model.* The predictions from the CNN and XGBoost models were used to train an ensemble model, as this approach has been successfully used to improve ML models in the past (Albuquerque et al., 2022; Meier et al., 2022). The linear combination of the *FVC* predictions from the CNN model ($\equiv FVC_{CNN}$) and XGBoost model ($\equiv FVC_{XGBoost}$) extracted from Fig. 4ac, is given by:

$$FVC_{ens} = w_1 \cdot FVC_{CNN} + w_2 \cdot FVC_{XGBoost} \tag{9}$$

where $w_1$ and $w_2$ are the coefficients of the linear combination and $FVC_{ens}$ is the final ensemble prediction. These coefficients are optimized by minimizing the error between $FVC_{ens}$ and the true *FVC* ($\equiv FVC_{true}$) for the training set. The loss function defined by Eq. (10) being minimized, $L$, is defined as:

$$L = \frac{1}{N}\sum_{i=1}^{N}\left|FVC_{true(i)} - FVC_{ens(i)}\right| \tag{10}$$

where the index $i$ refers to each of the $N$ samples of the training set.

## 3. Results and discussion

Fig. 3 shows representative LM images, exhibiting various *FVC*s. The flax fibers have a polygonal shape and are about 10 to 20 μm in diameter. Through the irregular hand-layup of flax fibers and the randomized locations in which images were taken, there is a broad range of local *FVC*s in the composite. This is ideal for training a CNN because the training set contains a wide variety of different input values. Another consequence of the composite preparation and image generation process is that the *FVC* shows an almost bell-shaped probability distribution function (see blue line in Fig. 3(b)). The minimum and maximum *FVC*s are about 5.9% and 54.8%, respectively. The mean and median *FVC*s are 28.3% and 28.6%, respectively, with a standard deviation of 8.8%. The distribution of *FVC* covers a wide range of values that are relevant for composites for load-bearing applications. The mean *FVC* is comparable to the *FVC* of other composites produced via hand-layup. The data set is representative of *FVC* that might be achieved in industrial applications and is suitable for the training of ML models (Gandhi et al., 2020; Mallick, 2007).

The CNN has 420,801 trainable parameters, and uses five 2D convolutional layers and three fully connected layers before the output layer (the optimized hyperparameters are shown in Fig. 2). The number of convolutional layers was also tuned by hyperparameter optimization (see Section 2). Fig. 4(a) shows the training error of the optimized CNN model, where the model was first trained with the training set and then used to predict *FVC* on the same training set. A small *MAE* error was obtained, as well as a good $R^2$ score. The same trained model still performed well on the test set, showing little overfitting, as the test error (2.89) was only a bit worse than the training error (1.76). Fig. 4(b) shows that the CNN model generalizes well and can reliably predict *FVC* using new LM images with a relatively small error. Note that the test set contains only original (non-augmented) images, which were never seen before by the trained CNN model.

The best hyperparameters found for the XGBoost model were: n_estimators = 100, learning_rate = 0.2, max_depth = 3, reg_alpha = 0.5, reg_lambda = 0.01. Even though regularization and a small tree depth were used, which are also supposed to decrease overfitting by decreasing the model complexity, the XGBoost model still exhibited some overfitting, as the training error shown in Fig. 4(c) is somehow smaller than the test error (Fig. 4(d)). Interestingly, the prediction performance on the test set is still acceptable, considering that XGBoost is not specialized in image recognition, that only every fifth pixel was used, and that many images were rotated or translated (via data augmentation), which makes more difficult to use flattened pixels for predicting *FVC*. The simplest model generally used as reference (least squares model) was not able to predict *FVC*.

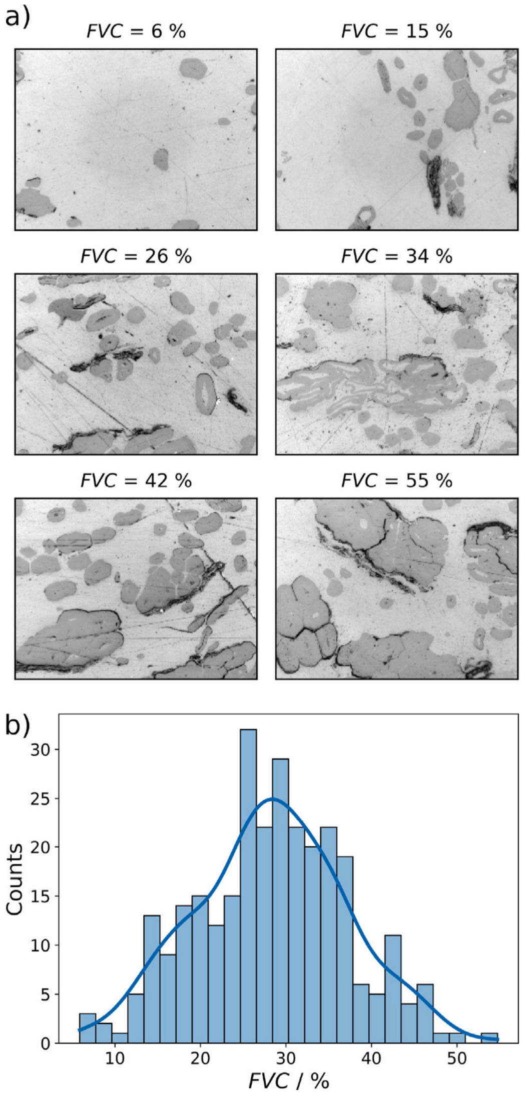

**Fig. 3.** (a) Representative LM images of the studied samples are depicted, showcasing various *FVC*s. The fibers, represented in dark gray, are oriented perpendicular to the plane of observation, while the epoxy matrix is represented in light gray. (b) The distribution of *FVC* values is presented in a bar graph, with the kernel density estimation illustrated as a blue line. These visualizations are derived from the analysis of 290 LM images employed in this study. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The full performance evaluation for both ML models for the training and test sets are compared in Table 1. The comparison of the *MSE* and *MAPE* errors for the training and test sets of both investigated ML models also confirms the overfitting trend discussed above. Both the XGBoost and CNN models have similar degree of overfitting, but the absolute errors are considerably smaller for the latter.

The optimized ensemble model had $w_1 = 0.671$ and $w_2 = 0.362$ as final parameters (see Eq. (9)) and showed a considerably better performance on the test set, as revealed in Fig. 5. For instance, the $R^2$

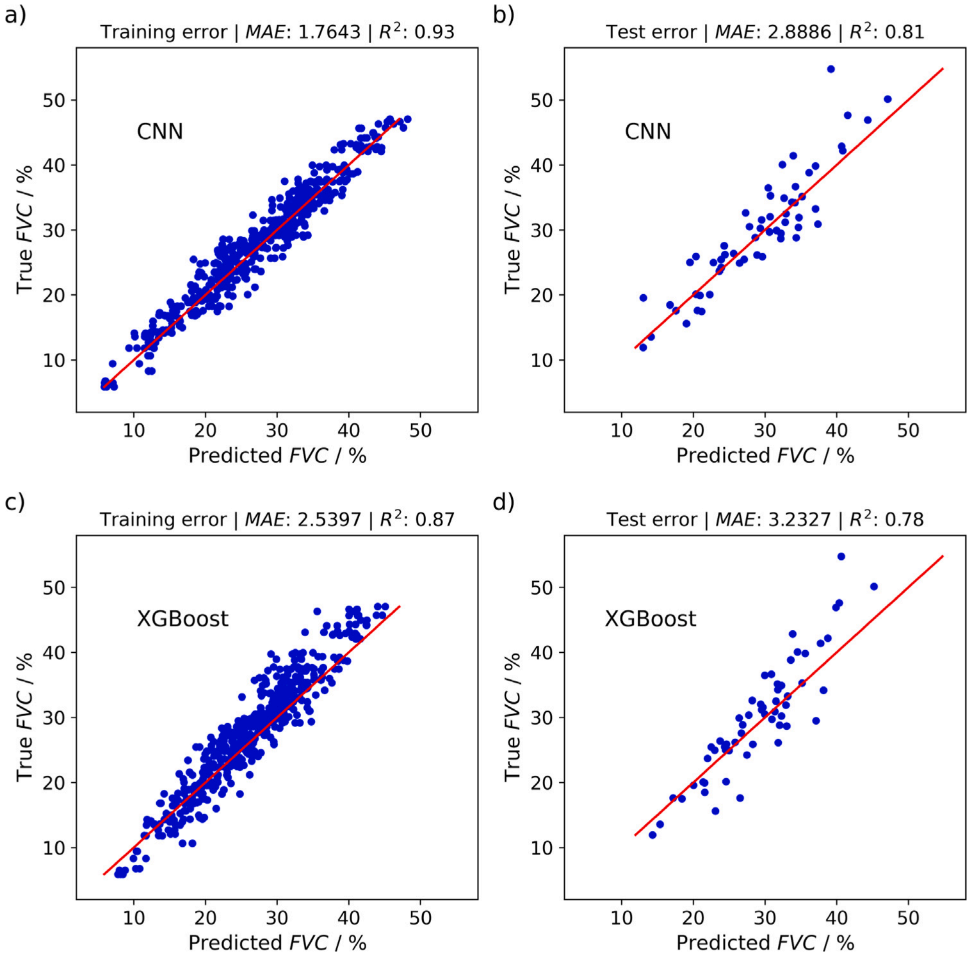

**Fig. 4.** Model evaluation using the training (2552 original + augmented images) and test (58 original images) sets employing the optimized CNN and XGBoost models for the prediction of *FVC*. Every 5th image from the training set is plotted for clarity.

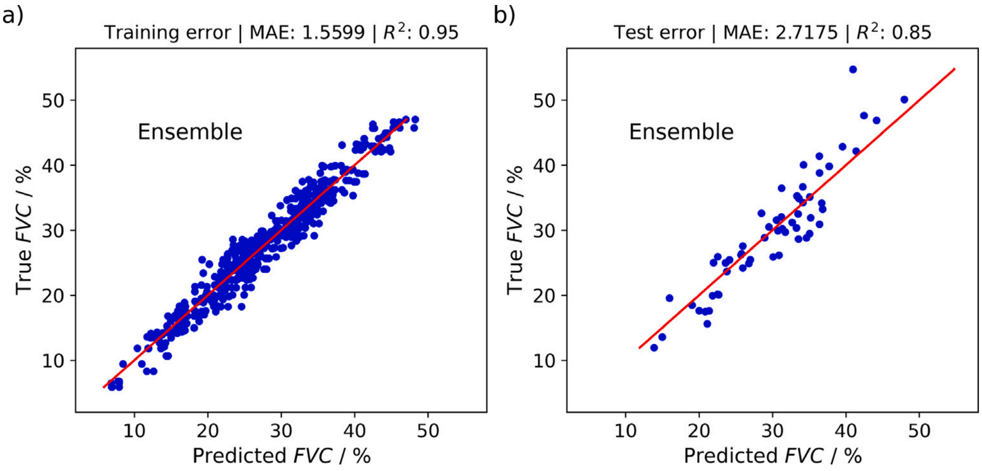

**Fig. 5.** Model evaluation using the training (2552 original + augmented images) and test (58 original images) sets employing the ensemble model (Eq. (9) with $w_1 = 0.671$ and $w_2 = 0.362$) for the prediction of *FVC*. Every 5th image from the training set is plotted for clarity.

score increased from 0.81 (CNN model) to 0.85 (ensemble model) for the test set. The larger value of $w_1$ means that predictions from the CNN model are considerably more important than that of the XGBoost to calculate the final ensemble prediction. This happens because the ensemble model (optimized with the training set) had much smaller MAE error for the CNN model compared to the XGBoost model. The full metrics calculated for the ensemble model are also shown in Table 1 for both the training and test sets, from where the superior performance of this model can be seen.

An error analysis has been performed for the ensemble model predictions carried out on the training and test sets since this model showed the highest accuracy. Fig. 6(a) shows histograms (vertical bars) of the predicted *FVC* for both the training and test sets using the ensemble model where each distribution has also its own kernel density

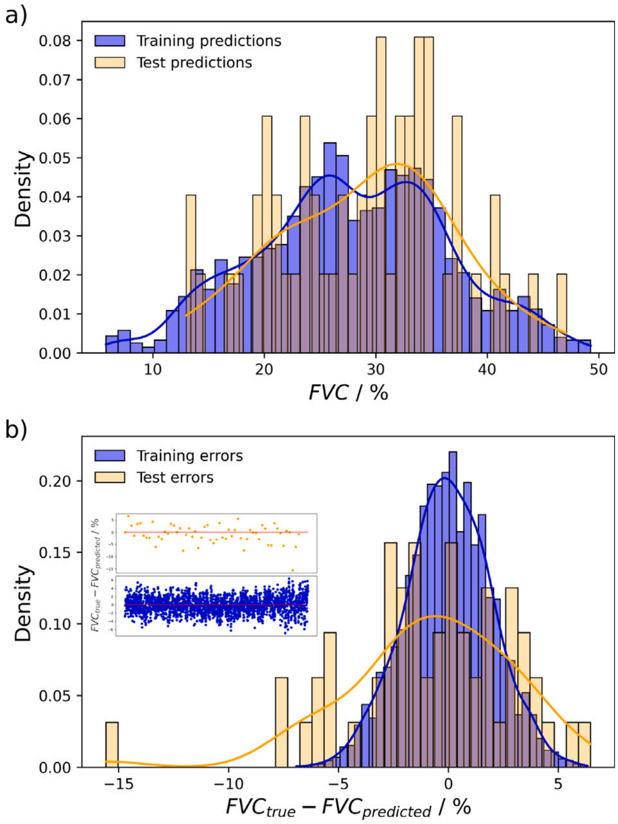

**Fig. 6.** (a) Histograms of ensemble predictions on the training and test sets (bars), each with its own Kernel Density Estimate (KDE, lines). (b) Histograms of the errors for the predictions shown in (a), each with its own KDE (lines). The inset shows the error scattering in each case.

**Table 1**
Model performance metrics for the investigated ML models trained with the training set with predictions done on the training (*tr*) and test (*te*) sets.

| Model | *MAE* | $R^2$ | *MAPE*/% | *MSE* |
|---|---|---|---|---|
| XGBoost $^{tr}$ | 2.54 | 0.87 | 9.69 | 10.04 |
| XGBoost $^{te}$ | 3.23 | 0.78 | 11.00 | 17.78 |
| CNN $^{tr}$ | 1.76 | 0.93 | 7.00 | 5.10 |
| CNN $^{te}$ | 2.89 | 0.81 | 9.63 | 15.14 |
| Ensemble $^{tr}$ | 1.56 | 0.95 | 6.32 | 3.85 |
| Ensemble $^{te}$ | 2.72 | 0.85 | 9.54 | 12.22 |

estimate (continuous lines). Fig. 6(a) shows that the distributions of training and test predictions nicely overlap and also have a similar shape to a Gaussian distribution with two peaks, which arises from the use of two ML models in the ensemble model. Fig. 6(b) shows that the errors of the predictions are rather unstructured (see the inset), being approximately described by a Gaussian distribution with zero mean. As expected, the Gaussian is broader for the test errors, as these predictions were performed on unseen data and are therefore supposed to exhibit larger errors.

The final accuracy found for the ensemble model (Table 1, last line) is already good enough for *FVC* characterizations required in polymer engineering, where acceptable *MAE* errors lie in the range of about 3.00–5.00. There are several ways of further improving the accuracy of the model. Re-parameterizing the CNN using Bayesian optimization would be one a first way. Adding other ML models to the ensemble approach would be a second way. Another way is to employ transfer learning to use one (or a combination of) the state-of-the-art deep learning models such as, for instance, those from the ResNet (He et al., 2016), VGGNet (Simonyan & Zisserman, 2015), and Inception Networks (Szegedy et al., 2016) families, and fine tune these models. AutoML libraries such as aX (Balandat et al., 2023) is still another way to do that. While this is certainly expected to improve even more the accuracy, it brings no new insight to the present proof-of-concept investigation and is out of the scope of this work. This is part of a future investigation (vide infra). Having compared the CNN discussed in this work with a much simpler ML model like XGBoost brings, on the other side, much more insight, as it provides alternatives to easily and efficiently train and optimize a ML model, therefore enhancing sustainability, as well as helping democratize knowledge since regions with low-budget computational resources could also have access to it.

To advance the discussion on the performance of ML models, it is crucial to explore both the potential and limitations of determining *FVC* in NFC using LM and CNN. Accurate analysis of *FVC* using LM-based methods hinges on two key assumptions: first, the fibers must be unidirectionally oriented. Second, the cross-sectional plane must be perpendicular to the fiber axes. Additionally, the quality of sample preparation is paramount, requiring polished surfaces free from scratches. Achieving this can be challenging due to the softness of natural fibers relative to glass or carbon fibers. Furthermore, sufficient contrast between the natural fibers and the polymeric matrix is necessary to enable effective binarization of the LM images via auto-thresholding.

Comparing ML models to existing techniques for determining the *FVC* in NFC is essential for objectively evaluating the models' practical usability. However, direct comparison between different experimental methods presents challenges. For instance, LM assesses the *FVC* of a microscopic section (200 μm by 150 μm) of the NFC, whereas TGA and density measurements require macroscopic samples for reliable *FVC* determination. It is important to note that, unlike LM, NFC samples analyzed by TGA or density measurements do not require unidirectional fiber orientation. However, applying these common methods to NFC remains complex due to the low thermal stability and low atomic density of natural fibers. The primary aim of this study is to demonstrate how image analysis of NFC via LM can be optimized with CNNs. Here, it is shown that automated image processing software, such as ImageJ, significantly reduces the time needed for *FVC* analysis when using macros compared to manual methods. Additionally, once CNNs and ensemble models are properly trained, they further expedite *FVC* assessment in LM images while maintaining robust predictive performance.

## 4. Conclusion

It has been proved that the pixels from light microscopy images of NFC could be reliably used to predict *FVC* using different ML models. The good accuracy achieved for both models and specially for their ensemble already fulfills the minimum accuracy requirements found in polymer engineering. The exhaustive maximization of the accuracy by using more complex deep learning models and transfer learning is out of the scope of this proof-of-concept investigation.

The effectiveness of CNNs and XGBoost models in predicting the *FVC* of NFC from LM images was shown. The distribution of *FVC*s, captured through a comprehensive set of LM images, exhibited a wide range, with values pertinent to industrial applications of load-bearing composites.

The CNN model, equipped with 432,301 trainable parameters and a series of convolutional and fully connected layers, showed good performance on both the training and test datasets. It achieved low errors and a high $R^2$ value on the training set, with little overfitting observed when evaluated against the test set. This suggests that the CNN model has a strong ability to generalize from the training data to unseen data, making it a reliable tool for predicting *FVC* in natural fiber composites based on microscopy images.

The XGBoost model displayed a similar degree of overfitting. It is noteworthy that the XGBoost model still delivered a good prediction accuracy on the test set, which was acceptably high for *FVC* characterizations required in polymer engineering. This outcome underscores the versatility of XGBoost in handling diverse datasets and learning tasks, even those for which it is not inherently specialized, like image recognition. Given that this model is easier, faster, and more efficient to optimize and train compared to CNN-based models, it presents an interesting alternative for predicting *FVC*. Not only is it more sustainable, but it also helps to democratize knowledge in regions with limited computational resources.

Ensembling the predictions from the CNN and XGBoost models enabled one to achieve even higher predictive performances for both the training and test sets.

The findings advocate for the integration of ML models into the analysis of NFC. The demonstrated capability of these models to accurately predict *FVC* from LM images can significantly streamline the material characterization process, providing a rapid and efficient means of assessing composite quality and performance. Future work may focus on maximizing the accuracy using complex deep learning models and transfer learning, exploring the incorporation of additional image features, and extending the methodology to other types of composites and material characterizations. This research paves the way for advanced material design and quality control through the application of ML technologies.

## CRediT authorship contribution statement

**Florian Rothenhäusler:** Conceptualization, Investigation, Methodology, Writing – original draft. **Rodrigo Queiroz Albuquerque:** Conceptualization, Machine learning modeling, Methodology, Writing – original draft. **Marcel Sticher:** Investigation. **Christopher Kuenneth:** Writing – review & editing. **Holger Ruckdaeschel:** Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

Parts of the research documented in this manuscript have been funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) within the research project “EcoPrepregs - Grundlagenforschung zur Klärung der Struktur-Eigenschaftsbeziehungen von Epoxidharzen und Fasern aus nachwachsenden Rohstoffen zur Anwendung in der Sekundärstruktur von Flugzeugen” (grant # 20E1907A, Germany). Parts of this work were funded by the “Bayerisches Staatsministerium für Wissenschaft und Kunst” (grant # F.2-M7426.10.2. 1/4/16, Germany). The authors want to thank Lukas Endner and Adwaith Sudhakaran for their support during the experiments. We would like to thank all colleagues of the work group “Resins & Composites” at the Department of Polymer Engineering for their support.

## Data availability

The trained CNN and XGBoost models, the dataset and a demonstrator code (Python, Jupyter Notebook) are here: “https://github.com/Polymer-Engineering-University-Bayreuth/FiberVolumeContent_ConvNet”, which is freely accessible in the public GitHub repository.

## References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., .... Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Abdalla, F., Megat Ahmad, M. M. H., Sapuan, S., & Sahari, B. (2008). Determination of volume fraction values of filament wound glass and carbon fiber reinforced composites. *ARPN Journal of Engineering and Applied Sciences*, *3*, 7–11.

Affdl, J. H., & Kardos, J. (1976). The Halpin-Tsai equations: a review. *Polymer Engineering and Science*, *16*(5), 344–352.

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining*.

Albuquerque, R. Q., Brütting, C., Standau, T., & Ruckdäschel, H. (2022). A machine learning investigation of low-density polylactide batch foams. *e-Polymers*, *22*, 318–331.

Albuquerque, R. Q., Sarhadi, A., Demleitner, M., Ruckdäschel, H., & Eder, M. A. (2024). Fatigue damage reconstruction in glass/epoxy composites via thermal analysis and machine learning: A theoretical study. *Composite Structures*, *331*, Article 117855.

Andrade, R. M. d., & Eduardo, A. C. (2011). Methodology for automatic process of the fired ceramic tile’s internal defect using IR images and artificial neural network. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, *33*, 67–73.

Ashby, M. (1992). *Materials selection in mechanical design* (1st ed.). Oxford: Butterworth-Heinemann.

Balandat, M., Bakshy, E., Marthur, L., Scott, S., Zhang, A., Karrer, B., Blei, D., & Polley, E. (2023). Ax: A python library for adaptive experimentation. Version 0.2.5.

Blarr, J., Kunze, P., Kresin, N., Liebig, W. V., Inal, K., & Weidenmann, K. A. (2024). Novel thresholding method and convolutional neural network for fiber volume content determination from 3D $\mu$CT images. *NDT & E International*, *144*, Article 103067.

Bledzki, A., & Gassan, J. (1997). Natural fibre reinforced plastics. In *Handbook of engineering polymeric materials*. CRC Press.

Bos, H. L. (2004). *The potential of flax fibres as reinforcement for composite materials* (Ph.D. thesis), University of Eindhoven, Netherlands.

Chauhan, R., Ghanshala, K. K., & Joshi, R. (2018). Convolutional neural network (CNN) for image detection and recognition. In *2018 first international conference on secure cyber computing and communication* (pp. 278–282).

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. CoRR abs/1603.02754.

Chin, W.-K., Liu, H.-T., & Lee, Y.-D. (1988). Effects of fiber length and orientation distribution on the elastic modulus of short fiber reinforced thermoplastics. *Polymer Composites*, *9*(1), 27–35.

Das, S. (2011). Life cycle assessment of carbon fiber-reinforced polymer composites. *The International Journal of Life Cycle Assessment*, *16*, 268–282.

DIN EN ISO 1172:1998-12 (1998). *Textilglasverstärkte kunststoffe - prepregs, formmassen und laminate - bestimmung des textilglas- und mineralfüllstoffgehalts*. Beuth Verlag.

Ehrenstein, G. W. (2006). *Faserverbund-kunststoffe*. Munich: Carl Hanser Verlag GmbH & Co. KG.

Flemming, M., Ziegmann, G., & Roth, S. (1995). *Faserverbundbauweisen - fasern und matrices*. Berlin: Springer.

Fu, S.-Y., & Lauke, B. (1996). Effects of fiber length and fiber orientation distributions on the tensile strength of short-fiber-reinforced polymers. *Composites Science and Technology*, *56*(10), 1179–1190.

Fu, S.-Y., & Lauke, B. (1998). The elastic modulus of misaligned short-fiber-reinforced polymers. *Composites Science and Technology*, *58*(3–4), 389–400.

Gandhi, U., Goris, S., Osswald, T. A., & Song, Y.-Y. (2020). *Discontinuous fiber-reinforced composites: fundamentals and applications*. Carl Hanser Verlag GmbH Co KG.

Grund, D., Orlishausen, M., & Taha, I. (2019). Determination of fiber volume fraction of carbon fiber-reinforced polymer using thermogravimetric methods. *Polymer Testing*, *75*.

Gupta, V., Mittal, R., Sharma, P., Mennig, G., & Wolters, J. (1989). Some studies on glass fiber-reinforced polypropylene. Part II: Mechanical properties and their dependence on fiber length, interfacial adhesion, and fiber dispersion. *Polymer Composites*, *10*(1), 16–27.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hessman, P. A., Riedel, T., Welschinger, F., Hornberger, K., & Böhlke, T. (2019). Microstructural analysis of short glass fiber reinforced thermoplastics based on x-ray micro-computed tomography. *Composites Science and Technology*, *183*, Article 107752.

ImageJ (2023). Getting started with simple visualization options in ImageJ. (Accessed 06 March 2024).

Jock, C. P. (1986). Quantitative optical microscopy fiber volume methods for composites. *Journal of Reinforced Plastics and Composites*, *5*(2), 110–119.

Kumar, S., Prasad, L., Bijlwan, P. P., & Yadav, A. (2022). Thermogravimetric analysis of lignocellulosic leaf-based fiber-reinforced thermosets polymer composites: an overview. *Biomass Conversion and Biorefinery*.

Machello, C., Baghaei, K. A., Bazli, M., Hadigheh, A., Rajabipour, A., Arashpour, M., Rad, H. M., & Hassanli, R. (2024). Tree-based machine learning approach to modelling tensile strength retention of Fibre Reinforced Polymer composites exposed to elevated temperatures. *Composites Part B (Engineering)*, *270*, Article 111132.

Mallick, P. K. (2007). *Fiber-reinforced composites: materials, manufacturing, and design*. CRC Press.

Meier, S., Albuquerque, R. Q., Demleitner, M., & Ruckdäschel, H. (2022). Modeling glass transition temperatures of epoxy systems: a machine learning study. *Journal of Materials Science*, *57*(29), 13991–14002.

Mwaikambo, L. Y., & Ansell, M. P. (2001). The determination of porosity and cellulose content of plant fibers by density methods. *Journal of Materials Science Letters*, *20*(23), 2095–2096.

Osa-Uwagboe, N., Udu, A. G., Ghalati, M. K., Silberschmidt, V. V., Aremu, A., Dong, H., & Demirci, E. (2024). A machine learning-enabled prediction of damage properties for fiber-reinforced polymer composites under out-of-plane loading. *Engineering Structures*, *308*, Article 117970.

Schlicht, S., Jaksch, A., & Drummer, D. (2022). Inline quality control through optical deep learning-based porosity determination for powder bed fusion of polymers. *Polymers*, *14*(5).

Schürmann, H. (2007). *Konstruieren mit faser-kunststoff-verbunden*. Berlin: Springer.

Shokrollahi, Y., Nikahd, M. M., Gholami, K., & Azamirad, G. (2023). Deep learning techniques for predicting stress fields in composite materials: A superior alternative to finite element analysis. *Journal of Composites Science*, *7*(8).

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Song, L., Li, X., Yang, Y., Zhu, X., Guo, Q., & Yang, H. (2018). Detection of micro-defects on metal screw surfaces based on deep convolutional neural networks. *Sensors*, *18*(11), 3709.

Sorour, S. S., Saleh, C. A., & Shazly, M. (2024). A review on machine learning implementation for predicting and optimizing the mechanical behaviour of laminated fiber-reinforced polymer composites. *Heliyon*.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818–2826).

Voigt, W. (1889). Ueber die Beziehung zwischen den beiden Elasticitätsconstanten isotroper Körper. *Annalen der Physik*, *274*(12), 573–587.